www.sonal-info.com

Mode d’emploi

Sonal version 2.0.80

Alex ALBER
Université F. Rabelais (Tours) / UMR C.I.T.E.R.E.S
Chercheur associé au Centre d’Etudes de l’Emploi

Table des matières

1 Avant-propos 9
1.1 Quelles sont les spécificités de Sonal ? 9
- Accompagner toutes les étapes de la recherche 9
1.2 Quand et pourquoi utiliser Sonal ? 10
- Gérer des bandes sonores ou des vidéos 10
- Délimiter des extraits sonores et les thématiser 12
- Retranscrire le contenu des extraits 14
- Marquer les passages les plus intéressants 15
- Construire une « base de données qualitatives » 15
- Filtrer les extraits par leurs caractéristiques 16
- Dépouiller statistiquement son corpus 18
- Autres usages possibles… 21
- Compatibilité 22
Quelques informations importantes : 22
- Le système d'enregistrement 22
- Les sauvegardes manuelles 22
- Logique fonctionnelle 23
- Le paramétrage des polices 23
2 Comprendre l'organisation du logiciel 24
2.1 Le système de fichiers 24
- Le fichier Corpus 24
- Codage des données dans le fichier Corpus 25
- Les fichiers satellites 26
- Organisation du fichier Rtr 27
2.2 Les bases de données virtuelles 29
- La base des entretiens 29
- Le dictionnaire des variables 30
- La base des extraits 30
2.3 Les chartes de codage à l'intérieur des retranscriptions 30
- Les balises de locuteur 30
- Les points de synchronisation (normaux ou dérivants) 31
- Les formats de pondération 32
- Les commentaires 33

Construire un corpus 35

3 Se lancer ! 36
3.1 Première étape : créer un répertoire de projet 36
- Où installer ce répertoire de projet ? 36
3.2 Créer un nouveau corpus 36
- Les étapes de la construction du corpus 38

4 Ajouter des documents à votre corpus 40
- Types de fichiers compatibles 40
- Possibilités d'encodage 41
4.1 Enregistrer un document sonore 41
4.2 Ajouter un document sonore 47
4.3 Ajouter une vidéo 48
4.4 Ajouter un fichier texte 49
4.5 Ajouter un fichier Rtr 51
4.6 Ajouter plusieurs documents simultanément 51
4.7 Concaténer des fichiers son 52

5 Gérer les documents de votre corpus 54
5.1 Renommer un document 54
5.2 Changer de format de document (allégez vos corpus) 55
5.4 Dupliquer une piste 57
5.5 Colorier les documents 59
5.6 Ajouter des séparateurs (créez des « familles ») 59
- Gérer les séparateurs 61
5.7 Trier vos documents 61
- Les fonctions de tri 61
5.8 Supprimer des documents 62

6 Construire la grille thématique 64
6.1 Ajouter une thématique 65
6.2 Modifier une thématique 66
6.3 Supprimer une thématique 66
6.4 Déplacer les thématiques 66
6.5 Recoder une thématique 67
6.6 Ajouter des séparateurs thématiques 67
6.7 Créer des thématiques en cours de retranscription 67

7 Construire et renseigner la base de données 69
7.1 Ajouter une variable 70
7.2 Renseigner une variable 71
7.3 Modifier une variable ou une modalité 72
7.4 Recoder des modalités 73
7.5 Déplacer les variables 75
7.6 Supprimer des variables 75
7.7 Explorer la base de données 76
7.8 Exporter la base de données 78

Encoder des documents 79

- Qu'appelle-t-on encodage dans Sonal ? 80

8 Coder des enregistrements audio/vidéo 81
- Généralités 81
- Contrôler de la lecture de la bande audio/vidéo dans la fenêtre de codage 84
8.1 Créer des extraits 85
- Généralités sur les extraits 85
- 1) Créer un extrait au point le plus avancé 86
- 2) Créer un extrait à la position du lecteur 87
- 3) Créer un extrait à la souris. 88
8.2 Editer, valider les extraits 89
- Rouvrir et fermer un extrait 89
- Rouvrir le dernier extrait 90
- Quand valider les extraits ? 90
- Propriétés de la zone d'édition 90
8.3 Définir manuellement les coordonnées d'un extrait 91
8.4 Affecter une ou des thématique(s) à un extrait 92
8.5 Ajouter des mots-clés (tags) aux extraits 93
8.6 Retranscrire le contenu des extraits (mode dictée) 94
- Paramétrer les raccourcis de locuteurs 95
- Le mode dictée 96
- Les options de la zone d'édition 101
- Insérer des commentaires dans la retranscription 105
8.7 Découper un extrait 105
- Découper les extraits sonores 105
- Couper les extraits audio-textuels (à un point de synchronisation) 105
- Découper les extraits audio-textuels (sans point de synchronisation) 106
8.8 Ajouter un point de synchronisation 106
- Trouver la position d'un mot dans l'enregistrement: 106
- Trouver la position dans le texte d'un passage sonore : 107
- Placer un point de synchronisation 108
- Placer une bordure d'extrait (point de partition) 108
8.9 Corriger un point de synchronisation 108
- Corriger un point manuellement 108
- Corriger un point au clavier 109
- Déplacer un point dans le texte 110
8.10 Supprimer un point de synchronisation 110
8.11 Supprimer tous les points de synchronisation 110
8.12 Fusionner deux extraits 111
8.13 Supprimer un extrait 112
8.14 Sauvegarder le codage 113
- Que se passe-t-il lorsqu'on enregistre ? 113
8.15 Travailler avec de la vidéo 115

9 Resynchroniser une retranscription 116

- Dans quels cas resynchroniser ? ... 116
9.1 Etape 1 : Importer la version sonore de l'entretien ... 116
9.2 Etape 2 : Importer le texte de la retranscription ... 116
9.3 Etape 3 : Reformater les balises de locuteurs ... 118
9.4 Convertir les timecodes en points de synchro (éventuellement) ... 118
9.5 Etape 4 : premier découpage grossier ... 119
9.6 La synchronisation semi-automatique ... 120
- Ajouter des points dérivants ... 120
- Corriger les points dérivants ... 121

10 Encoder un document texte ... 123
- Spécificité du codage des documents texte ... 123
10.1 La page de codage du texte ... 124
- Utilisation ... 125
10.2 Découper des extraits dans les documents textuels ... 125
10.3 Affecter une thématique à un extrait textuel ... 126
10.4 Ajouter un nouvel extrait ... 126

11 La pondération qualitative ... 127
11.1 Marquer un passage ... 127
11.2 Règles à connaître ... 128
11.3 Modifier/annuler une pondération ... 130
- Changer le niveau de pondération ou la supprimer ... 130
- Redéfinir la fin de la partie pondérée ... 130

12 Les mots-clés (tags) ... 131
12.1 A quoi servent les mots-clés ? ... 131
12.2 Ajouter des mots-clés ... 131
12.3 Recoder des mots clés ... 132
12.4 Filtrer les extraits par mots clés ... 133

13 Exporter les retranscriptions ... 134
13.1 Les fichiers Rtr (bon à savoir) ... 134
13.2 Exporter une retranscription ... 134
- Ouvrir dans Word ... 135
- Exporter dans fichier Rtf ... 136
- Exporter en html ... 136
- Options d'affichage de l'export ... 136

Analyser un corpus 138

14 Organisation générale du système d'analyse..... 139
14.1 Choisir (ou pas) un angle d'approche des données 139
14.2 Le système de filtrage 140
- Filtrer par thématique 142
- Filtrer par entretien 143
- Filtrer par variables 143
- Filtrer les extraits par leurs mots-clés. 144

15 L'analyse qualitative (synthèse)..... 145
15.1 Modifier les extraits depuis la synthèse..... 146
15.2 Rechercher un mot clé 147
15.3 Filtrer les extraits par le « poids » de leur contenu 147
15.4 Exporter le résultat d'une synthèse 149
- Exporter le texte 149
- Exporter le son 149
15.5 Enregistrer une synthèse..... 150
15.6 Ouvrir une synthèse enregistrée..... 150

16 L'analyse lexicométrique 151
16.1 A quoi sert l'analyse lexicométrique? 151
- Compter les occurrences des différents mots (ou "formes") 151
- Etudier les spécificités langagières associées à un sujet et/ou à un (type de) locuteur etc. 151
- Etudier les catégories de répondants ou les sujets dans lesquels une forme est la plus fréquente. 152
16.2 La base lexicale 152
- Réinitialiser la base lexicale 153
16.3 La page de lexicométrie 154
- Les panneaux de filtrage 154
16.4 Le tri des mots 155
- Le tri par occurrences 155
- Le tri par ordre alphabétique..... 156
- Le tri par pourcentage des occurrences 156
- Le tri par spécificité..... 157
- Régler le seuil de spécificité..... 158
- Afficher (ou non) les spécificités négatives 158
- Filtrer la liste des formes 159
- Désactiver l'affichage instantané 159
- Inclure les questions dans l'analyse lexicale..... 159
16.5 Revenir au contexte d'apparition d'une forme..... 159
- Analyser un mot en détail..... 160
16.6 La lemmatisation 165
- La lemmatisation manuelle : « regrouper » 165
- Définir les mots outils 166
- La lemmatisation automatique..... 166

- Enregistrer les lemmes créés ... 167
- Recharger les lemmes créés ... 167
16.7 Gérer les groupes créés ... 168
16.8 L'analyse factorielle ... 168

17 L'analyse chronométrique ... 171
17.1 A quoi sert l'analyse chronométrique? ... 171
- Avertissement préalable : encoder n'est pas prouver ... 171
17.2 Réaliser une analyse chronométrique dans Sonal ... 171
17.3 Approfondir l'analyse : exporter au format *.Xls ou Trideux ... 173
17.4 Travailler des données chronométriques dans Trideux ... 174
17.5 (se) Lancer (dans) Trideux ... 176

18 L'analyse des échanges ... 180
18.1 Le tableau des interventions ... 180
18.2 l'analyse graphique des interventions ... 183

19 L'analyse tabulaire ... 185

Annexes ... 186

20 Utiliser Sonal dans une recherche collective ... 187
20.1 Créer un corpus modèle ... 188
- Construisez la grille thématique ... 188
- Renseignez le dictionnaire des variables ... 189
20.2 Distribuer le corpus modèle ... 191
20.3 Centraliser les fichiers ... 191
- Comment réintégrer un fichier externe au corpus central? ... 191

21 Débuguer un corpus ... 192
21.1 Diagnostiquer l'origine du problème ... 192
- Vérifier le fichier corpus ... 192
21.2 Si un entretien dysfonctionne ... 193
- Restaurer une version antérieure d'un entretien ... 193
- Réparer manuellement un entretien ... 194

22 Traduire Sonal ... 195
22.1 Sonal pour musiciens ... 195
22.2 Personnaliser une traduction ... 195

23 Informations ... 197

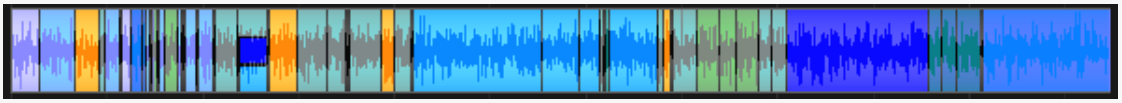
1 Avant-propos

Sonal est un logiciel de ***construction* et d'*analyse* de corpus de données qualitatives**. Vous pourrez l'utiliser pour **archiver, encoder, annoter, retranscrire puis analyser des documents sonores**, des **vidéos** et des **documents textuels**.

1.1 Quelles sont les spécificités de Sonal ?

Sonal a été conçu initialement pour répondre aux besoins spécifiques de son concepteur. Il s'agissait de mieux organiser un travail d'enquête qualitative par entretiens. Le logiciel s'est ensuite enrichi des apports des utilisateurs et des équipes de recherche qui l'ont utilisé. L'idée initiale était de **partir des matériaux primaires de la recherche (enregistrements sonores ou vidéo[1]),** et de **coder directement les bandes sonores**, au lieu d'attendre de retranscrire pour coder du texte, comme on le fait généralement. L'objectif était d'accompagner la recherche dès ses premières étapes.

L'originalité de Sonal est qu'il permet de découper **directement dans le son** des extraits, symbolisés, sur la représentation graphique des bandes, par des cubes auxquels on peut apposer une couleur correspondant à une thématique :



C'est à l'intérieur de ces extraits que seront ajoutées les retranscriptions. Le travail de retranscription n'intervient donc plus comme un moyen de remplacer la bande par un verbatim mais comme un moyen d'**ajouter du texte « *sur* » la bande**.

Le résultat du travail d'encodage/retranscription est donc un **matériel audio-textuel dit « synchronisé »** car le son et sa retranscription sont toujours accessibles simultanément. Sonal propose toute une série d'outils très spécifiques qui permettent de gérer conjointement les versions sonores et textuelles des entretiens (outils de resynchronisation, découpage/fusion simultanés du texte et du son etc.).

- Accompagner toutes les étapes de la recherche

En intégrant la gestion des entretiens dans leur version sonore, Sonal accompagne tout le spectre du travail d'enquête, de la collecte des matériaux bruts (il est possible d'enregistrer les entretiens directement dans Sonal) à leur analyse quantitative en passant par leur encodage et leur retranscription. **Sonal est donc un tout-en-un qui**

[1] Il est également possible de traiter des documents textuels dans Sonal, mais la méthode proposée est fortement marquée par la logique initiale du logiciel.

réunit plusieurs fonctions : celles d'un logiciel de gestion d'enregistrements, celles d'un logiciel de retranscription, et certaines fonctions clés que partagent la plupart des Caqdas[2].

Cette organisation contribue à rompre la frontière existant généralement entre la phase de retranscription (totale ou partielle) et la phase d'encodage et d'analyse. A l'inverse de ce qu'il se produit généralement, la version sonore des entretiens n'est plus délaissée au profit du texte après retranscription ; le retour à la bande est constamment possible, même en phase d'analyse. Cela introduit une grande souplesse dans l'organisation du travail. Les phases habituelles de l'enquête sont confondues et réversibles : **on peut encoder les bandes sans les retranscrire, retranscrire les bandes sans les encoder, retranscrire des entretiens préalablement encodés, encoder des entretiens déjà retranscrits. Le plus pratique étant cependant d'encoder en retranscrivant ou de retranscrire en encodant, ce qui constitue l'un des avantages du logiciel.** Dans tous les cas, rien n'oblige à commencer par une tâche donnée, toutes peuvent être réalisées à n'importe quel moment.

Il faut également savoir que quelles que soient l'avancée et l'organisation du travail, les outils d'analyse peuvent toujours être utilisés sur les matériaux tels qu'ils sont, pour fournir de premiers retours réflexifs. On peut **analyser des encodages sans retranscription, des retranscriptions sans encodage,** le mieux étant bien sûr d'analyser des entretiens intégralement retranscrits et encodés.

Sonal est donc d'abord un **outil d'assistance à la construction des corpus**, qui couvre un spectre de tâches assez large, tout en permettant une organisation plus libre des différentes étapes du travail.

1.2 Quand et pourquoi utiliser Sonal ?

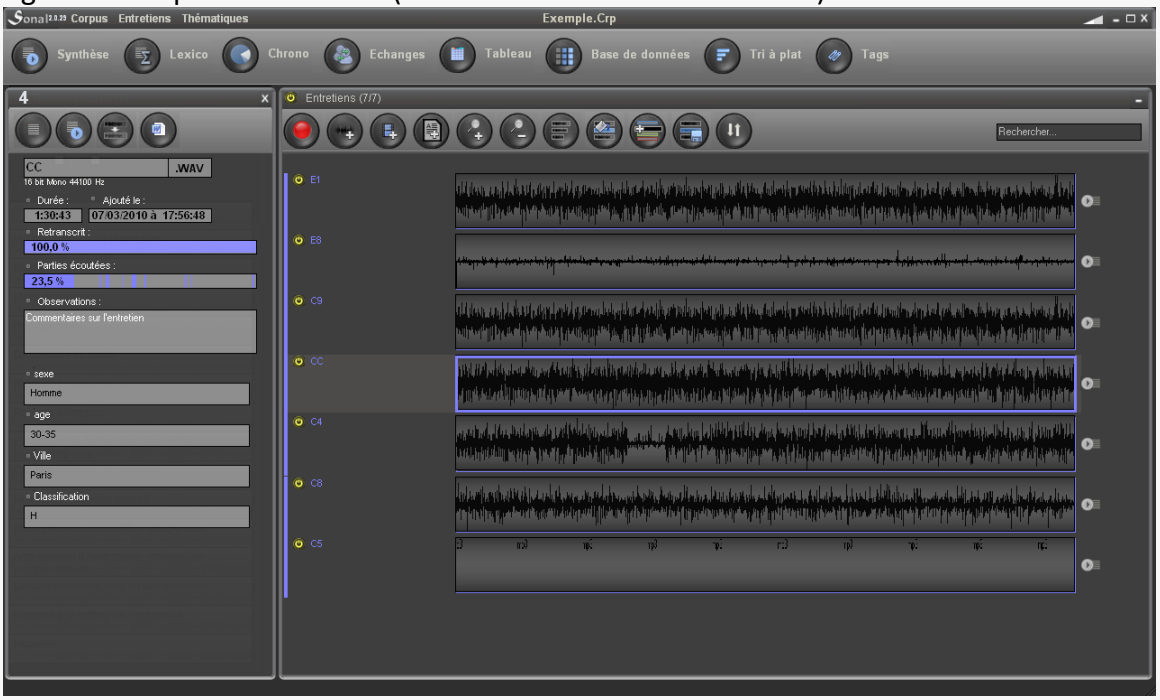
Faisons le point sur ce que vous pouvez attendre de Sonal.

- Gérer des bandes sonores ou des vidéos

La première fonction du logiciel est de vous permettre de gérer une collection de bandes sonores et/ou de vidéos : un « corpus » de fichiers textes et/ou multimédias (999 au maximum) qui pourront être commentés et décrits par l'intermédiaire de variables stockées dans une base de données.

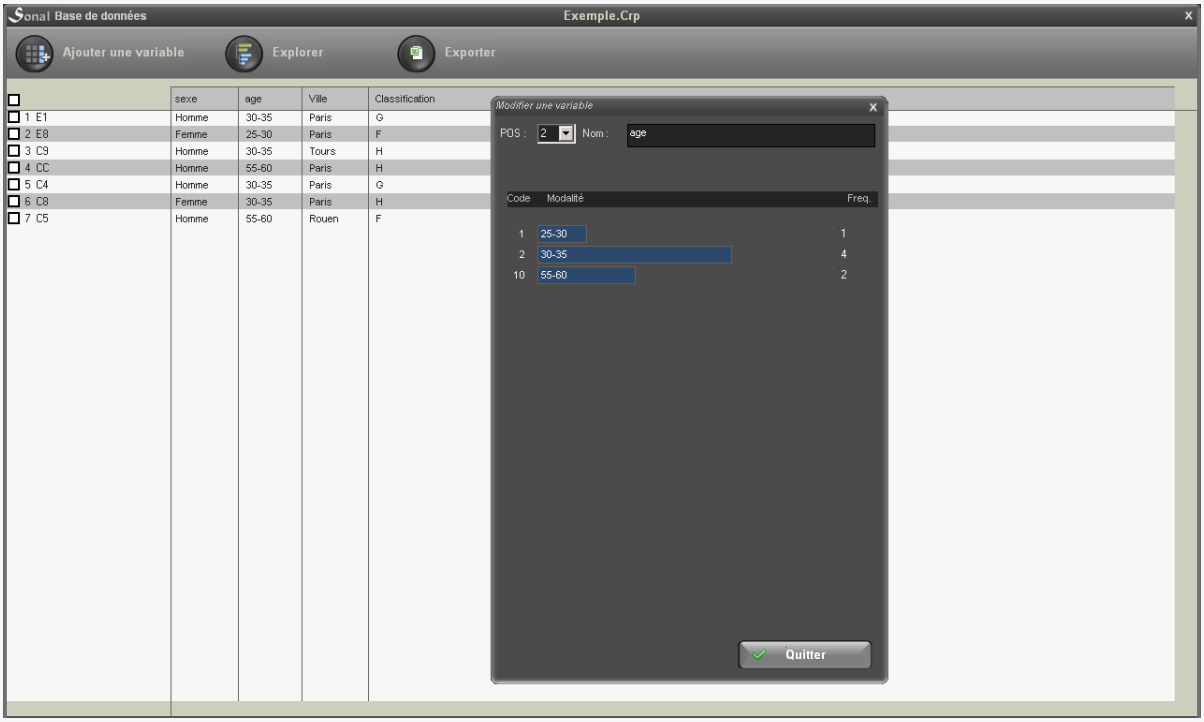
[2] Computer aided qualitative data analysis software

Figure 1 - Corpus d'entretiens (non retranscrits et non encodés)



Cette base de données (construite par l'utilisateur) contiendra, pour chaque entretien, une série de caractéristiques définies par l'utilisateur, sous la forme de variables (99 au maximum) pouvant prendre chacune jusqu'à 98 modalités (99= non réponse). Vous aurez accès à un décompte des différentes caractéristiques des fichiers (tri à plat). La base de données, comme les analyses de fréquence, est exportable vers n'importe quel tableur.

Figure 2 – Base de données Sonal (avec tri à plat d'une variable)

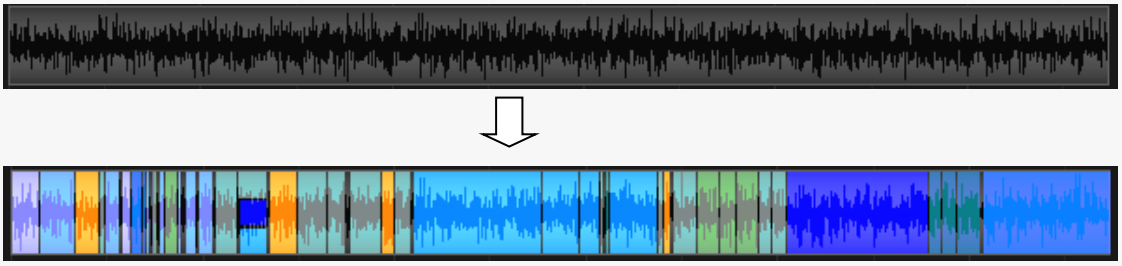


La base de données permet aussi de filtrer et trier les entretiens selon leurs caractéristiques (type d'enquêtés, type de documents…) lors de la phase d'analyse ou tout simplement pour faciliter la recherche d'un entretien spécifique dans un corpus de grande taille.

- Délimiter des extraits sonores et les thématiser

Vous pouvez encoder chaque fichier pris individuellement en délimitant des portions de bande homogènes (c'est-à-dire des parties de discours traitant d'un même sujet ou ayant des caractéristiques communes) par l'apposition d'un marqueur coloré directement sur le spectre graphique des entretiens. Il s'agit, pour faire simple, d'utiliser une sorte de feutre surligneur directement sur les fichiers media (son et vidéo) pour repérer certaines parties. Ces parties de bande ainsi isolées et coloriées sont appelées des « extraits ».

Figure 3 - Exemple de fichier son avant et après encodage

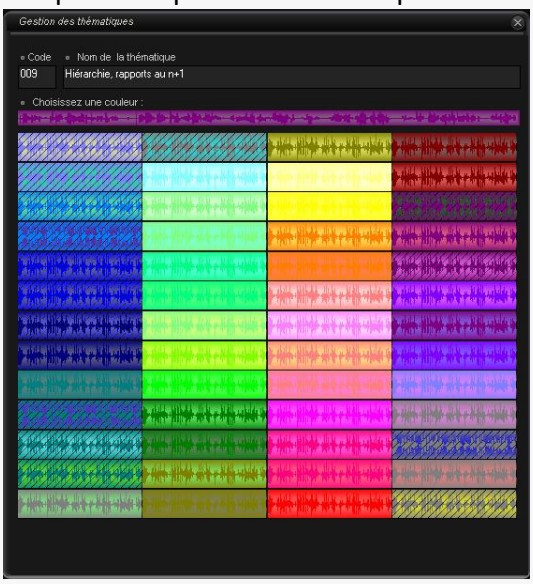


Les couleurs apposées sur la bande sont appelées « thématiques » car leur usage premier est de repérer des sujets particuliers dans les entretiens, mais de nombreux usages détournés sont possibles. Par exemple, dans la version de Sonal pour musiciens, les thématiques sont appelées « Styles ».

L'utilisateur peut créer jusqu'à 99 thématiques, distinguées par une couleur, et en affecter une ou plusieurs à chacun des extraits.

Figure 4 – Fenêtre de gestion des thématiques

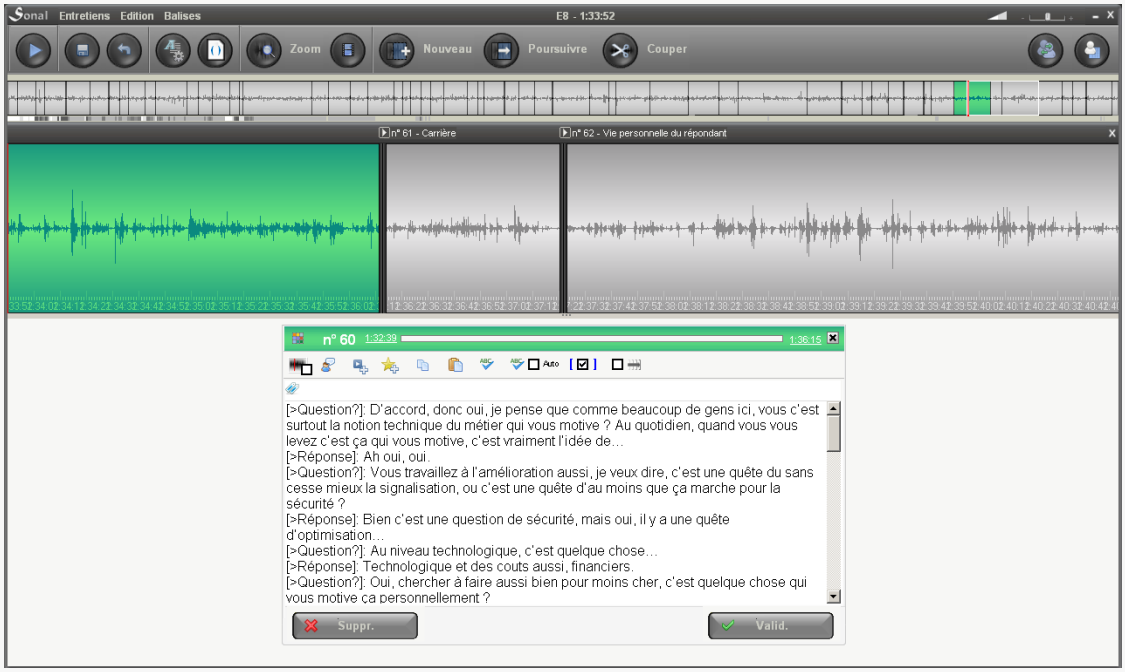
Figure 5 – Palette de couleurs disponibles pour les thématiques



- Retranscrire le contenu des extraits

Une fois que la bande sonore est découpée en extraits, il est possible d'ouvrir chacun d'entre eux pour y ajouter du texte.

Figure 6 – Interface de codage/retranscription



On peut d'abord placer à l'intérieur des zones de texte une simple description du contenu, des notes ou - et c'est l'orientation première de l'outil - y **retranscrire le contenu des prises de parole.**

Sonal assiste l'utilisateur dans cette tâche à l'aide d'un **système de dictée** qui fonctionne ainsi : la lecture progresse par courtes portions successives (durée déterminée par l'utilisateur), s'arrête et attend la validation du transcripteur pour passer à la portion suivante. La bande progresse donc au rythme du transcripteur, sans que celui-ci ait à s'en occuper outre mesure. De plus, Sonal laisse périodiquement des balises dans le texte afin de renseigner la position des passages retranscrits (appelés « points de synchronisation »). Enfin, un système de balises de locuteurs annonçant les prises de parole (raccourcis clavier F1 à F12 paramétrables) permet de distinguer les questions des réponses.

Si l'encodage est continu (tous les extraits se touchent) et que l'intégralité des extraits est retranscrite, les versions sonores et textuelles de l'entretien sont dites « synchronisées ». On peut dorénavant naviguer soit dans le son, et relire immédiatement la retranscription qui en a été faite, soit évoluer dans la retranscription, et retrouver très rapidement le passage sonore associé à ce qu'on vient de lire.

- Marquer les passages les plus intéressants

Un système de « pondération qualitative » permet de formater de manière spécifique certains passages du texte retranscrit ; concrètement, de mettre en gras et d'agrandir la police de certains passages (5 niveaux d'agrandissement sont proposés) afin de repérer des passages intéressants.

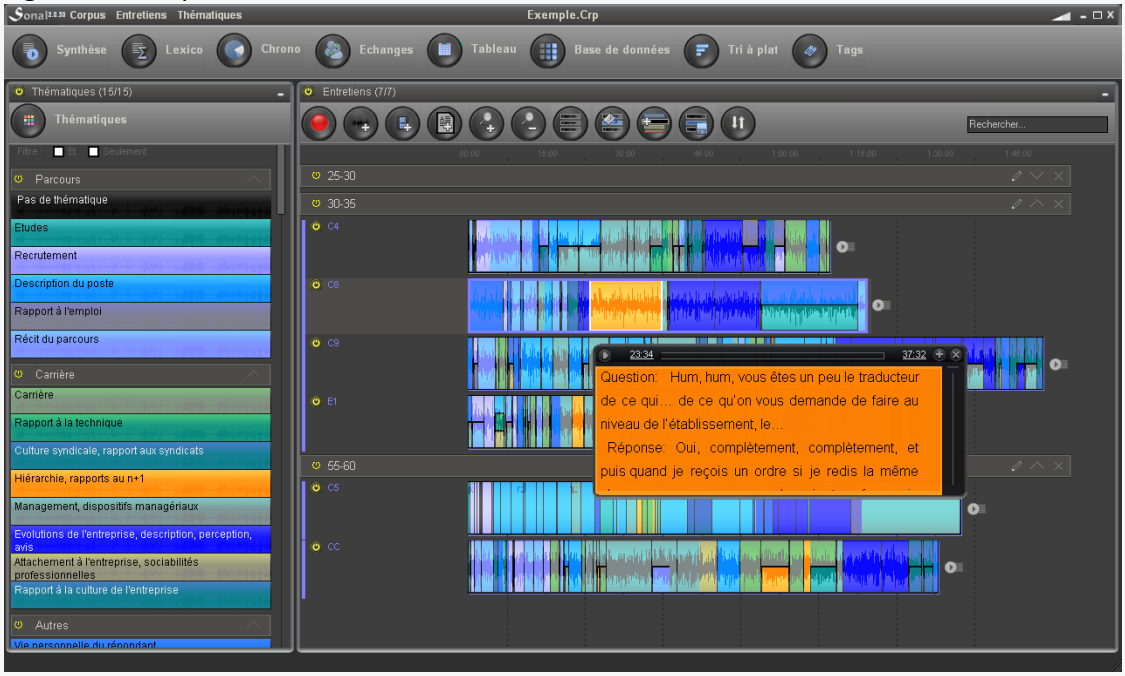
La différence avec ce même travail réalisé dans un traitement de texte est que **ce formatage intervient ensuite comme critère de filtrage** à l'intérieur des extraits préalablement isolés : seuls les passages marqués au niveau donné apparaîtront. On peut ainsi condenser des passages entiers en les résumant par les quelques phrases les plus révélatrices ou les plus édifiantes au regard de notre grille d'analyse.

- Construire une « base de données qualitatives »

Lorsque tous les entretiens d'un corpus ont été encodés, ce dernier se transforme en une sorte de « base de données qualitatives » composée d'une collection d'extraits possédant chacun une ou plusieurs série de caractéristiques : des coordonnées temporelles, une ou plusieurs thématiques, un contenu textuel, une série de mots-

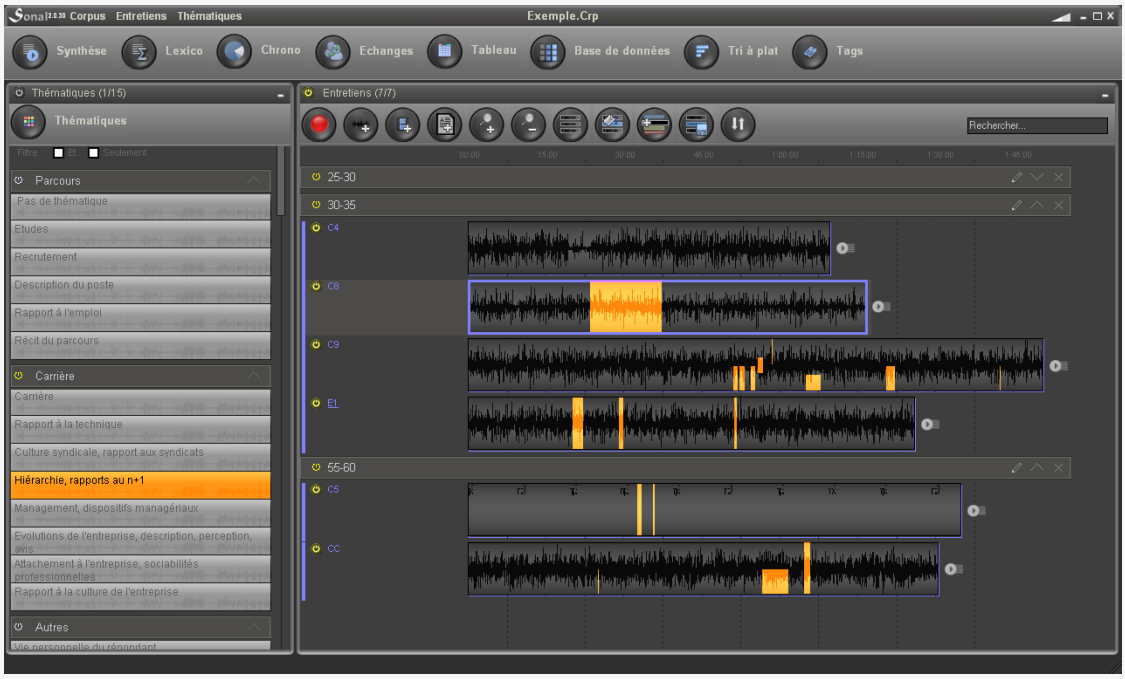
clés. Chaque extrait possède également les caractéristiques attribuées dans la base de données à l'entretien auquel il appartient. Tous les entretiens se superposent dans une sorte de « mur » dans lequel chaque brique colorée correspond à un extrait.

Figure 7 – Corpus de démonstration, dont un extrait est en cours d'écoute

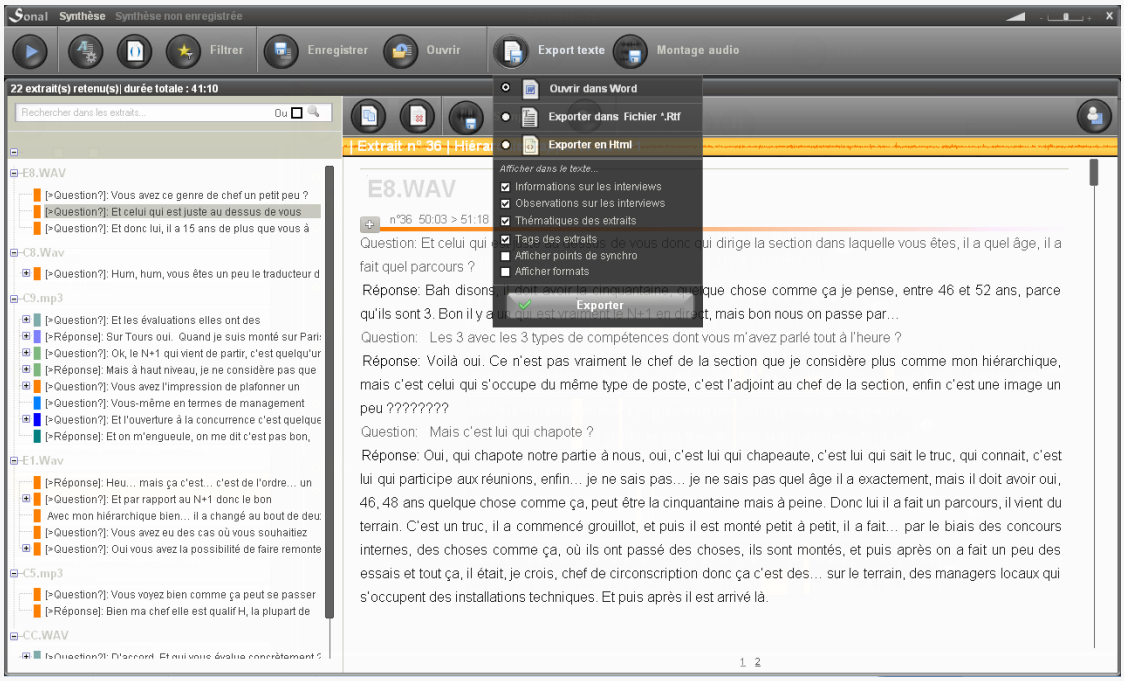


- Filtrer les extraits par leurs caractéristiques

Le fait d'avoir ainsi une collection d'extraits dotés de caractéristiques directement tirées du cadre d'analyse utilisé dans la recherche (variables, thématiques) permet d'opérer des filtrages très utiles pour éprouver ses hypothèses. Chaque caractéristique des extraits peut en effet servir d'opérateur de filtrage : sa ou ses thématiques, son contenu textuel, les mots-clés et enfin les caractéristiques de l'entretien duquel il est issu peuvent être utilisés de manière simultanée pour isoler tous les extraits répondant aux caractéristiques recherchées. La sélection se fait suivant un système d'allumage/extinction des extraits qui isole certaines « briques » du mur d'entretiens.



On retrouve ainsi très rapidement tous les extraits pertinents, qui sont envoyés dans une fenêtre de « synthèse ». Les synthèses peuvent être enregistrées dans Sonal, et être exportés vers un traitement de texte pour insertion dans une restitution écrite.

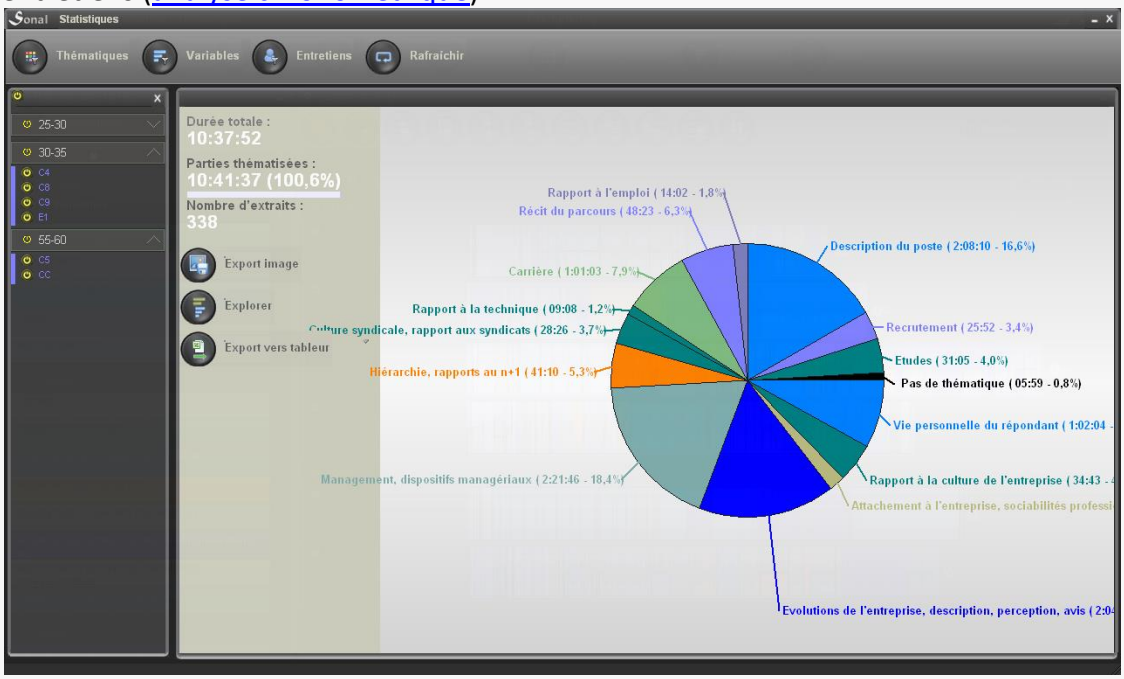


En permettant un inventaire rapide des matériaux disponibles sur les différents sujets, ce système répond aux questions prosaïques : qu'ai-je comme matériau sur telle ou telle thématique ? Qu'en disent les répondants ayant tel ou tel profil ? Selon le niveau de traitement des matériaux, le résultat sera sous la forme d'un micro-trottoir (son uniquement) ou d'un thésaurus de passages audio-textuels.

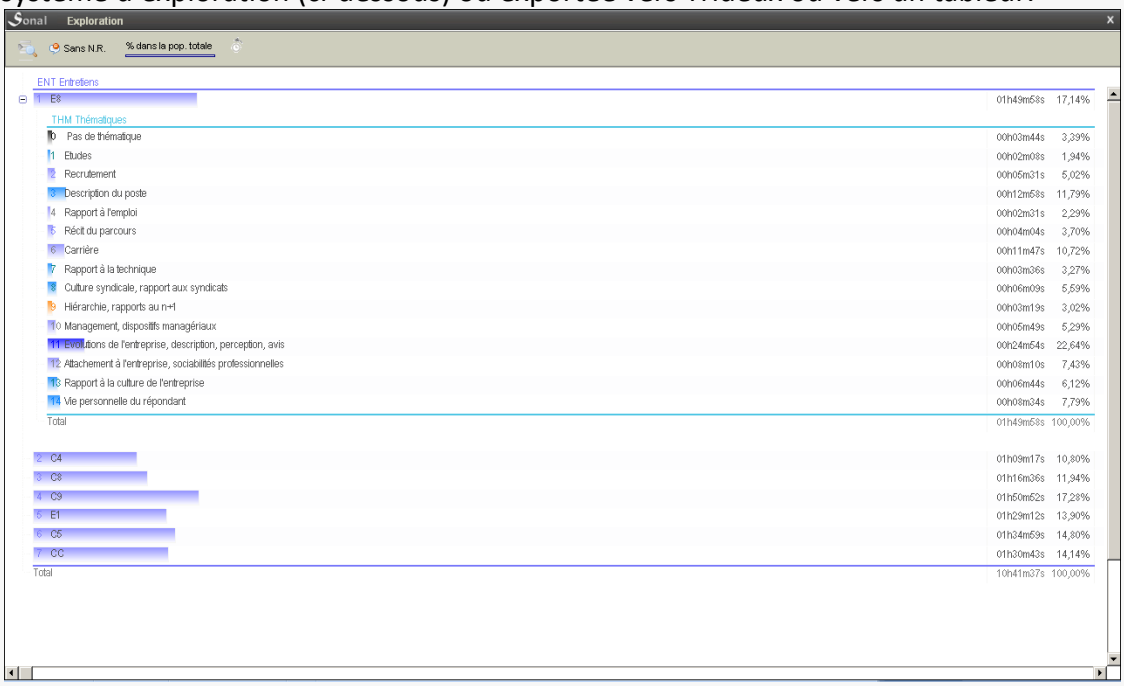
- Dépouiller statistiquement son corpus

Pour sortir d'une analyse purement compréhensive, Sonal propose ensuite d'opérer des dénombrements sur les matériaux du corpus. Deux dimensions différentes peuvent être quantifiées : le temps ou le texte, soit la durée ou le contenu des extraits.

On peut par exemple comparer le temps consacré aux différents sujets dans les entretiens (analyse chronométrique).



Cette analyse chronométrique peut ensuite être analysée en profondeur via un système d'exploration (ci-dessous) ou exportée vers Trideux ou vers un tableur.

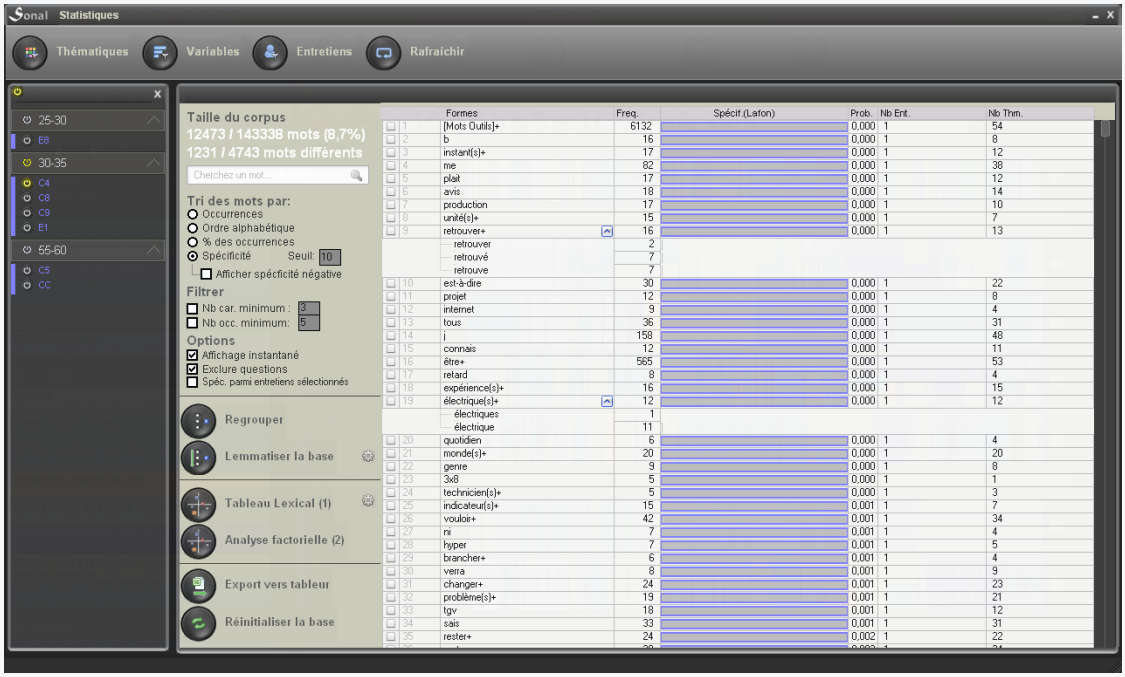


Sonal Exploration

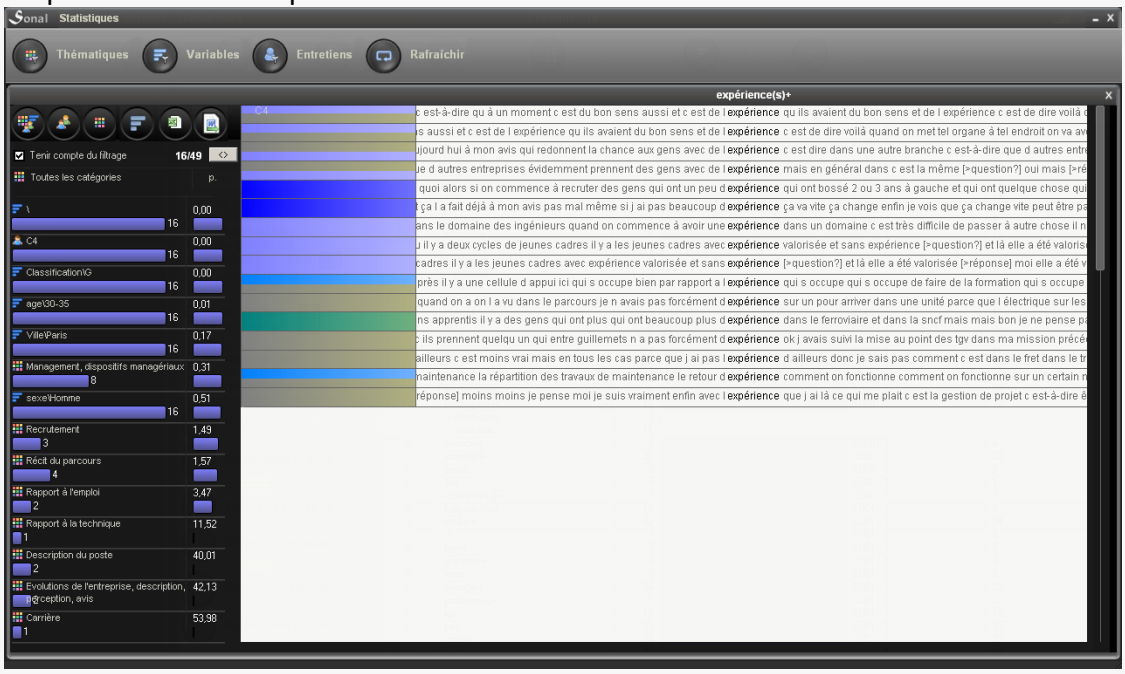
Sans N.R. | % dans la pop. totale

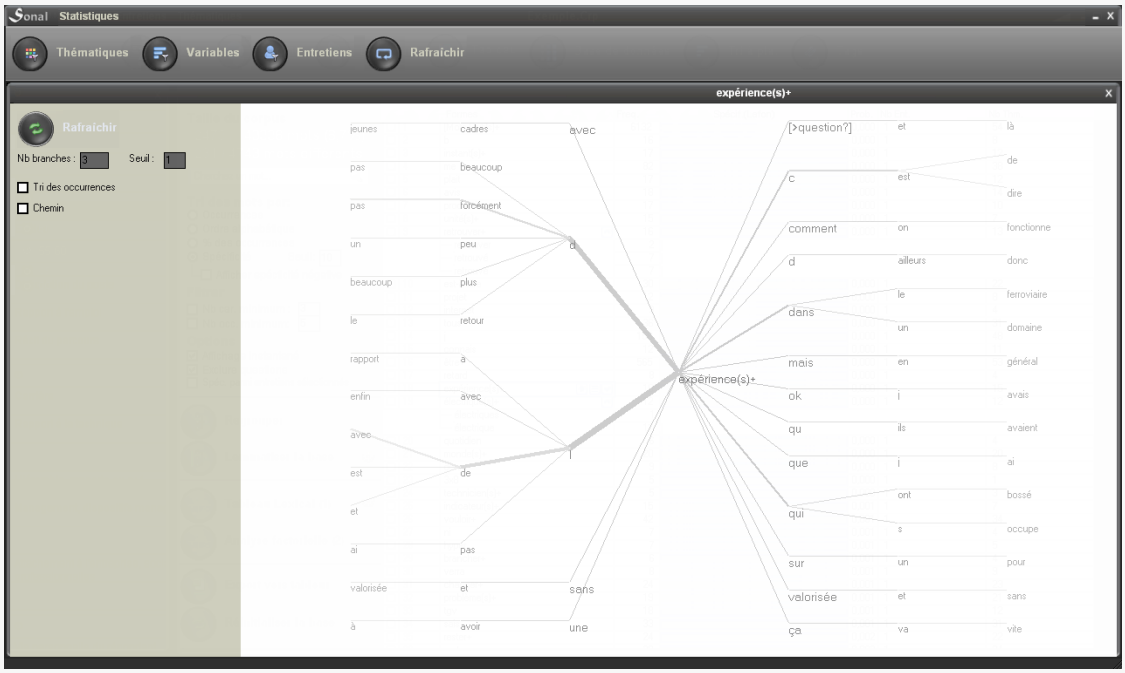
ENT Entretiens			
1 E8		01h49m58s	17,14%
THM Thématiques			
0	Pas de thématique	00h03m44s	3,39%
1	Etudes	00h02m08s	1,94%
2	Recrutement	00h05m31s	5,02%
3	Description du poste	00h12m58s	11,79%
4	Rapport à l'emploi	00h02m31s	2,29%
5	Récit du parcours	00h04m04s	3,70%
6	Carrière	00h11m47s	10,72%
7	Rapport à la technique	00h03m36s	3,27%
8	Culture syndicale, rapport aux syndicats	00h06m09s	5,59%
9	Hiérarchie, rapports au n+1	00h03m19s	3,02%
10	Management, dispositifs managériaux	00h05m49s	5,29%
11	Evolutions de l'entreprise, description, perception, avis	00h24m54s	22,64%
12	Attachement à l'entreprise, sociabilités professionnelles	00h08m10s	7,43%
13	Rapport à la culture de l'entreprise	00h06m44s	6,12%
14	Vie personnelle du répondant	00h08m34s	7,79%
Total		01h49m58s	100,00%
2 C4		01h09m17s	10,80%
3 C8		01h16m36s	11,94%
4 C9		01h50m52s	17,28%
5 E1		01h29m12s	13,90%
6 C5		01h34m59s	14,80%
7 CC		01h30m43s	14,14%
Total		10h41m37s	100,00%

On peut également analyser le contenu textuel des extraits (analyse lexicométrique), par exemple repérer les mots les plus spécifiquement utilisés selon les enregistrements, les types d'enregistrement, les sujets abordés, les locuteurs etc.

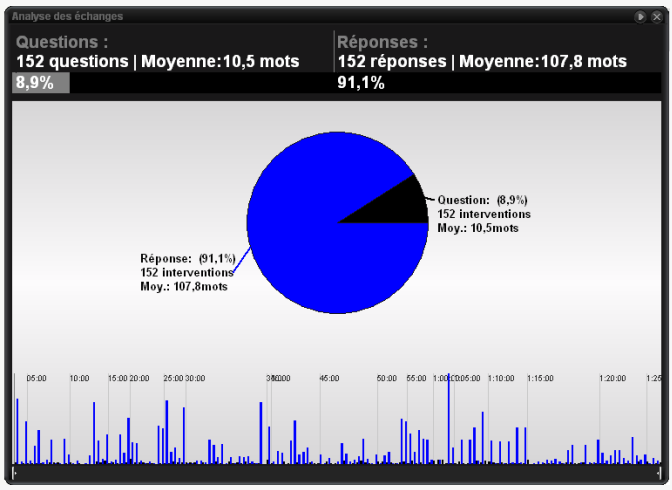


Une lemmatisation manuelle ou automatique des formes (intégralement réversible) est disponible, qui permet de créer des entités sémantiques de niveau supérieur.
On peut également étudier un mot en particulier, pour savoir dans quelles parties du corpus il apparaît le plus significativement, et se pencher sur les formes dans lesquelles il revient le plus souvent.

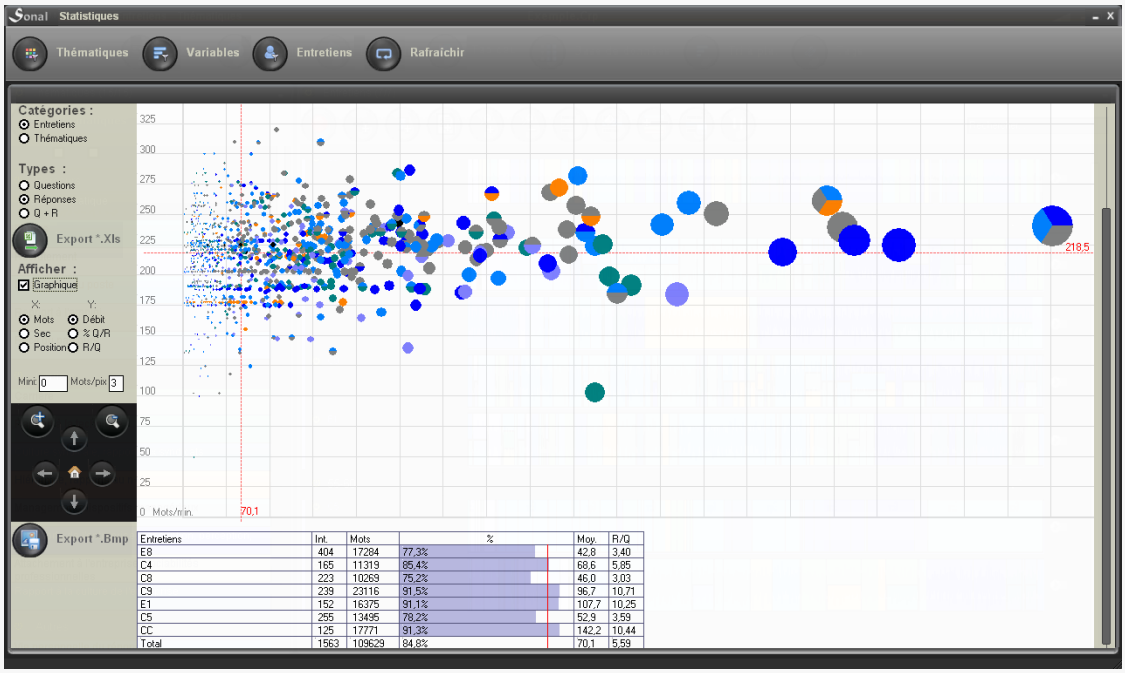




Il est également possible d'étudier la dynamique des conversations, en connaissant, pour chaque entretien, le ratio des questions et des réponses, la longueur moyenne des interventions etc.



Enfin, le croisement des contenus sonores (durée) et textuels (nombre de mot) des extrait débouche sur une analyse de débit (débitmétrie) qui analyse les variations de la vitesse d'élocution en fonction des répondants, des sujets etc.)



On peut donc à la fois avoir une approche compréhensive, consistant à étudier successivement des sélections d'extraits pour synthétiser leur contenu, ou utiliser les fonctions d'analyse globale pour se laisser suggérer des pistes analytiques dont on pourra très facilement tester la pertinence en retournant aux matériaux d'origine.

- Autres usages possibles…

Si Sonal a d'abord été pensé pour constituer des corpus d'entretiens sociologiques, on peut envisager de nombreux usages plus ou moins détournés, en fonction de la nature des enregistrements audio et/ou vidéo intégrés et des textes dans le corpus. Dès lors que l'on a des bandes sonores ou des vidéos que l'on souhaite archiver, on peut utiliser s'en servir.

On peut par exemple compiler des cours enregistrés, en repérant les différentes sections et les chapitres, enregistrer des séminaires de recherche ou des congrès et repérer les différentes prises de paroles, des questions etc. Moins académiquement, Sonal peut - pourquoi pas - permettre à un groupe de musique d'enregistrer ses répétitions en intégralité, pour repérer certaines improvisions intéressantes, ou certains passages à retravailler. De même pour toute production visuelle ou sonore de longue durée que l'on souhaite annoter. Enfin, il est également possible d'utiliser Sonal pour réaliser des **sous-titrages**. Les points de synchronisation permettent en effet de repérer la position temporelle des extraits dans la bande, et sonal les affiche dans la fenêtre vidéo de manière synchronisée, ce qui revient à transformer les retranscriptions en sous-titres. .

- Compatibilité

Sonal est un logiciel développé dans et pour un environnement Windows (XP à 8). Il est cependant possible de l'utiliser dans un environnement Mac ou Linux, mais cela suppose une émulation de Windows.

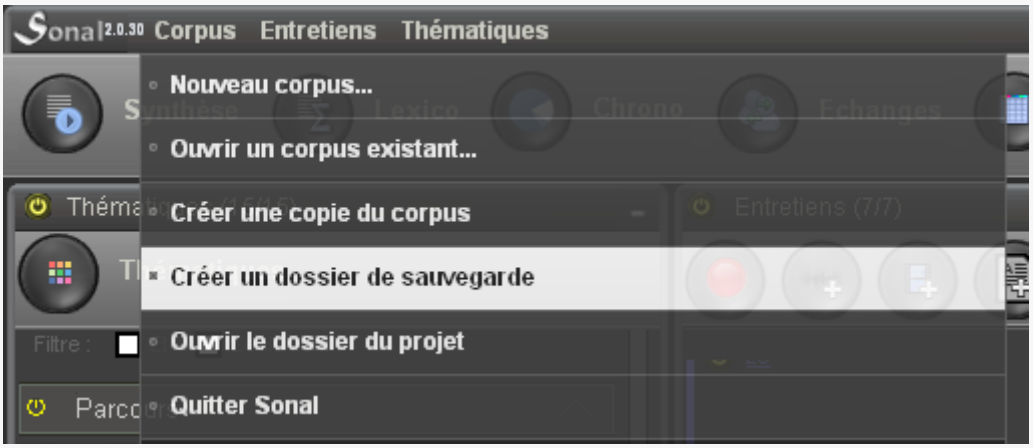
Quelques informations importantes :

- Le système d'enregistrement

Sonal a un fonctionnement en **sauvegarde continue**. Il part du principe que toutes les modifications que vous faites doivent être sauvegardées. Il n'y a donc peu de boutons « enregistrer ». **L'enregistrement est continu, par exemple, à chaque fois que vous retranscrivez une portion de dictée, que vous redécoupez des extraits, que vous les recoloriez etc. Sonal enregistre vos actions. Tout est fait pour limiter le risque de perte de données.**

- Les sauvegardes manuelles

Pour plus de sécurité, réalisez fréquemment des sauvegardes de votre travail. Dès que vous avez un peu progressé dans la saisie de vos entretiens ou dans le renseignement de la base de données, créez un dossier de sauvegarde depuis la page principale.



C'est une copie des fichiers essentiels utilisés par Sonal (cf. ci-dessous) qui sont placés dans un dossier à part.

> *Les fichiers média ne sont pas recopiés, vous pouvez donc en créer très fréquemment sans prendre le risque de surcharger votre ordinateur. Par exemple, la sauvegarde de 15h d'enregistrements intégralement retranscrits ne pèse même pas un mégaoctet.*

Vous pouvez même vous envoyer ce dossier compressé par mail pour plus de sécurité. *Il vaut donc mieux faire trop de sauvegardes que pas assez.*

- Logique fonctionnelle

- Pour changer quelque chose, cliquez dessus. Si une zone de texte apparaît, c'est que vous avez trouvé. Pour enregistrer la modification d'un paramètre, faites entrée.

- Il n'y a pas de doubles-clics dans le logiciel.

- Ne faites qu'une chose à la fois. En plus de la page du corpus, éviter d'ouvrir plusieurs pages conjointement (synthèse en même temps que saisie etc.).

- Le paramétrage des polices

Sonal est un logiciel un peu particulier en cela qu'il s'adapte à la taille de l'écran en dehors des procédures prévues par Windows. Cela le rend sensible à un problème particulier :

Si votre définition d'écran est élevée, il est possible que Windows vous ait proposé d'agrandir les polices de 25% : Sonal n'aime pas ça du tout ! Les fenêtres sont mal dimensionnées et vous ne verrez pas toutes les fonctions. Vous devez travailler avec des polices de taille normale.

2 Comprendre l'organisation du logiciel

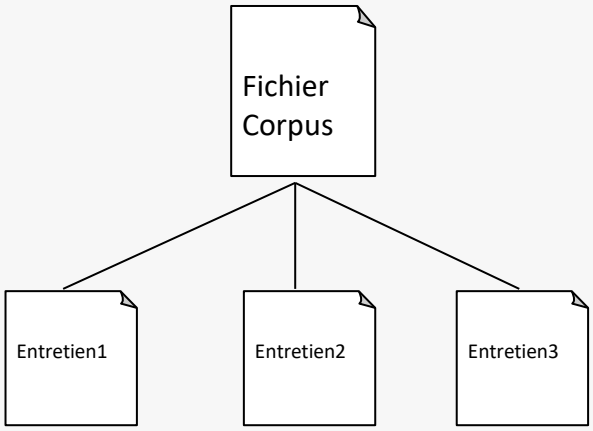
Cette partie vise à vous permettre de comprendre la manière dont Sonal fonctionne, ce qui vous aidera à corriger d'éventuels problèmes.

2.1 Le système de fichiers

Sonal a un fonctionnement en adresses relatives. Il ne mémorise pas la position des différents fichiers qu'il utilise, seulement leur nom, car il connaît leur position par leur présence dans un seul et même dossier. Ce fonctionnement oblige donc à mettre **tous les fichiers d'une enquête dans un seul et même dossier**.

> *Cela a quelques inconvénients mais beaucoup d'avantages, notamment celui de vous permettre de transférer facilement vos projets d'un ordinateur à un autre, de rouvrir facilement un corpus depuis un support de sauvegarde etc.*

A l'intérieur du répertoire projet, il y a deux types de fichiers différents : un fichier central, le **fichier projet** (corpus) dont le rôle est de centraliser les informations communes à tous les fichiers d'un projet, et les **fichiers satellites**, qui ne se réfèrent qu'à un seul entretien.



- Le fichier Corpus

Le fichier Corpus a pour extension *.crp. C'est en ouvrant un fichier corpus que l'on accède aux entretiens retranscrits d'un projet.

Le fichier corpus ne contient aucune donnée sonore ni aucune retranscription. Sa fonction est de réunir les informations communes à tous les entretiens du projet.

Les informations stockées dans le fichier corpus servent ensuite à interpréter le contenu des fichiers satellites rattachés aux entretiens.

> *Le fichier d'encodage (*.rtr) d'un entretien donné ne contiendra pas l'information « le répondant est un homme de 30 à 40 ans » mais, « à la variable 1, le répondant a la modalité 1 et à la variable 2, le répondant a la modalité 3 », en suivant les valeurs de l'exemple ci-dessus. De même, un extrait de cet entretien ne se verra pas affecter une thématique en toutes lettres (comme par exemple « Thématique 2 »), mais uniquement le code de la thématique (« 004 » dans l'exemple ci-dessus).*

Ce fonctionnement a des incidences particulières pour tout ce qui a trait aux recodages, ce pourquoi il vaut mieux opérer toutes les modifications depuis le logiciel lui-même, qui va aller appliquer les modifications (dans le code d'une variable ou d'une thématique par exemple) à la fois dans le fichier corpus et dans tous les fichiers satellites.

- Codage des données dans le fichier Corpus

C'est un simple fichier texte (éditable avec le bloc note) qui contient, entre balises :

- La liste des thématiques : |**THEM**|

> *Chaque **thématique** est décrite par des informations séparées par des virgules :*
> *1) le nom court (ex : 001)*
> *2) la couleur (valeur hexadécimale. Ex : 16744576)*
> *3) Le nom long (ex : thématique 1)*
>
> *Les **groupes thématiques** sont déclarés par un intercalaire annoncé par un « > » suivi du nom du groupe (ex : Groupe1). Toutes les thématiques qui suivent cet intercalaire lui sont affiliés, jusqu'à l'intercalaire suivant ou la fin de la liste.*

- La liste des variables de la base de données : |**POS**|

> *Chaque variable est déclarée sur une ligne par :*
> *1) une position (entre 1 et 99)*
> *2) un nom long (exemple : « Sexe », « Age » etc.)*
> *3) un statut (public =1 ou privé= 0)*

- Le dictionnaire des modalités des différentes variables : |**DIC**|

> *Chaque modalité est déclarée sur une ligne par :*
> *1) une variable (désignée par sa position)*
> *2) une modalité (entre 1 et 99)*
> *3) un intitulé (exemple : « Homme »)*

- La liste des entretiens |**ENTR**|

> *Chaque entretien est déclaré sur une ligne par :*
> *1) Le nom du fichier*
> *2) la couleur de l'entretien (valeur hexadécimale. Ex : 65280)*
> *3) Si l'entretien est actif ou non (1 ou 0)*

- La liste des locuteurs répondant aux codes F1 à F12 **|LOCU|**

Exemple de contenu d'un fichier corpus :

```
Projet crée le 09/10/2012 à 19:10:48
<|THEM|
 000,&H0, Pas de thématique
 >Groupe1
 001,16744576,Thématique 1
 004,16744448,Thématique 2
 >Groupe2
 005,16744448,Thématique 3
 007,16744448,Thématique 4
>|THEM|
<|POS|
 1,Sexe
 2,Age
>|POS|
<|DIC|
 1,1,Homme
 1,2,Femme
 2,1,- de 20 ans
 2,2,20-29 ans
 2,3,30-39 ans
 2,4,40-49 ans
>|DIC|
<|ENTR|
 Entretien 1.Wav,65280,1
 Entretien 2.Wav,9342606,1
 Document1.Rtr, 65280,0
>|ENTR|
<|LOCU|
 Question?
 Réponse
 R2
 R3
 R4
 R5
 R6
 R7
 R8
 R9
 R10
 R11
>|LOCU|
```

- Les fichiers satellites

On a pu remarquer que Sonal ne stockait que le nom des bandes contenues dans le projet (« entretien1.wav » etc.). Pourtant, chaque fichier qui est intégré dans un corpus Sonal va faire l'objet d'un encodage et/ou d'une retranscription. Il faut donc que d'autres fichiers soient créés en parallèle du fichier son. La logique de fonctionnement est assez simple : **tous les fichiers créés en lien avec un entretien ou**

un enregistrement donné portent le même nom que lui, et ne s'en différencient que par leur extension.

Pour chaque élément de la collection (entretien, bande, vidéo), il y a **plusieurs fichiers différents**, qui portent le même nom mais se différencient par leur extension :

- Un fichier **media** (audio, vidéo) ou un **fichier Texte.** C'est ce fichier dont le nom est stocké dans le fichier corpus entre les balises |ENTR|.

> *Attention, lorsque sont ajoutés des documents sans bande son, c'est directement le nom du fichier rtr (voir ci-dessous) qui est renseigné dans la liste des entretiens du fichier crp.*

- Un fichier *.**Bmp** : c'est l'**image du spectre sonore** de l'intégralité de l'entretien. Pour gagner du temps, il n'est dessiné qu'une fois lors de l'ajout de l'entretien au projet et enregistré.

- Un fichier *.**Lct** : c'est le fichier qui **mémorise les parties lues**.

- Un fichier *.**Rtr** : c'est le fichier qui contient les **informations sur l'entretien** (durée, observations, modalités prises par les différentes variables) et surtout **les parties retranscrites** et leurs coordonnées.

Le fichier RTR CONTIENT TOUT VOTRE TRAVAIL. C'est le fichier le plus important de tous.

- Organisation du fichier Rtr

La première ligne contient la durée en secondes.

Puis, les balises|OBS| délimitent l'espace laissé aux observations au sujet de l'entretien. Les balises |ATTR| délimitent l'espace laissé aux attributs. Sur chaque ligne, une valeur de variable puis de modalité, séparées par une virgule. C'est le fichier corpus qui permettra d'interpréter ces codes (voir ci-dessus).

Ensuite, sont listés les différents extraits repérés dans la bande.

Chaque extrait contient un **entête** qui renseigne d'abord la position de début en secondes, entre deux « :: ». Exemple ::310,55 :: signifie que l'extrait débute à 5 minutes, 10 secondes et 55 centièmes.

Cette position de départ est suivie par la ou les thématiques affectée(s) à l'extrait, toujours sous la forme « 000 ». S'il y a plusieurs thématiques, celles-ci sont séparées par une virgule. Enfin, suite à « Tags= », les mots clés éventuels affectés à l'extrait, séparés par une virgule.

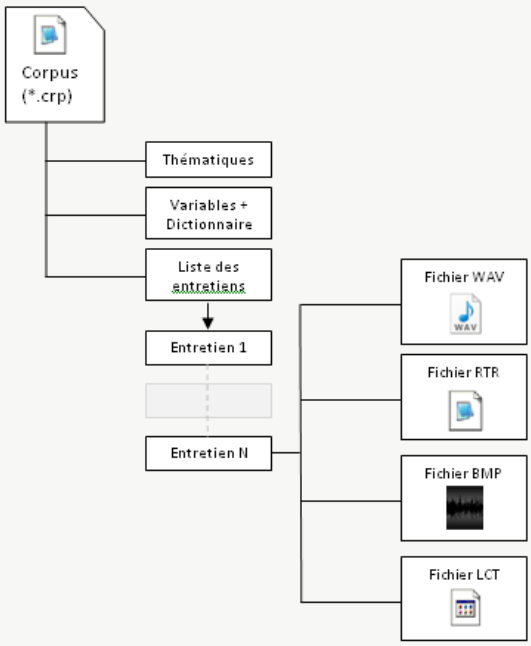
Sous cette ligne d'entête, le texte de l'extrait.

Enfin, après un saut de ligne, les coordonnées de la fin de l’extrait, suivant la même charte que pour l’entête.

Exemple de contenu d’un fichier RTR :

```
46
<|OBS|
Observations laissées au sujet de l’entretien
>|OBS|
<|ATTR|
1,1
2,3
>|ATTR|
::0,00::001 Tags=intéressant, à réécouter
Texte associé au premier extrait
::10,00::
::10,00::001,004 Tags=digression,
Texte associé au deuxième extrait
::17,00::
::17,00::000 Tags=
Texte associé au troisième extrait
::27,00::
::27,00::000 Tags=
Texte associé au quatrième (et dernier) extrait
::45,00::
```

Organisation générale des fichiers



> *Malgré ses contraintes, ce fonctionnement permet une certaine souplesse. Pour peu que leurs grilles thématiques et leurs variables soient identiques, deux fichiers Crp peuvent largement « partager » un même fichier RTR. On peut aussi décliner un corpus en plusieurs sous-corpus. Par exemple, un dossier de projet pourra contenir un corpus complet, une copie qui ne contient que certains types d'entretiens, une autre que les entretiens d'un interrogateur donné etc. en se référant tous à des fichiers communs. Ce qui impliquera que le travail dans un entretien s'impactera dans tous les projets où il est intégré.*

2.2 Les bases de données virtuelles

Lorsqu'un corpus est chargé, Sonal lit successivement les fichiers RTR de tous les entretiens d'un corpus et crée 3 bases de données en mémoire virtuelle.

- La base des entretiens

Cette première base renseigne, pour chaque entretien :

- son nom
- sa durée
- les observations qui lui sont rattachées
- son nombre d'extraits
- les valeurs associées à chacune des variables.
- sa couleur

- Le dictionnaire des variables

Parallèlement, un dictionnaire est créé, à partir du contenu du fichier corpus. Il renseigne, pour chaque variable, l'intitulé de toutes les modalités renseignées. Les limites sont fixées préalablement à 99 variables pouvant prendre chacune 98 modalités.

- La base des extraits

La troisième base est un tableau de données renseignant, pour chaque extrait de chaque entretien :

- ses coordonnées
- la ou les thématiques
- le texte associé
- les mots-clés.

Le dimensionnement du tableau de données est fixé par défaut à 999 extraits dans 999 entretiens. Normalement, cela suffit amplement.

Le croisement de ces bases permet ensuite à Sonal d'opérer toutes sortes d'analyses résultant de filtrages opérés tant sur le contenu des extraits que sur les caractéristiques des entretiens desquels ils sont issus. Lorsque Sonal ouvre un extrait, il sait par la base des extraits dans quel entretien il se trouve. A l'aide de la base des entretiens, il sait quelles sont les caractéristiques de cet entretien et à l'aide du dictionnaire, il peut renvoyer si, par exemple, l'extrait est issu d'un entretien « homme » ou « femme », et s'il faut l'afficher ou non, en fonction du filtrage opéré.

2.3 Les chartes de codage à l'intérieur des retranscriptions

Sonal permet d'enrichir le texte des retranscriptions d'informations utiles pour l'analyse ou le travail de codage. Il s'agit toujours de balises repérées dans le texte par une forme spécifique et représentées différemment dans la version lisible du texte. Ces formats sont spécifiques à Sonal. Ils sont simples et peuvent être saisis ou corrigés manuellement dans le texte des retranscriptions. Pour une utilisation complète de Sonal, il importe de les connaître et de comprendre leur fonctionnement.

- Les balises de locuteur

Les balises de locuteurs servent à repérer les tours de parole dans les documents. Outre qu'elles simplifient la lecture (les questions sont grisées par exemple), elles sont très utiles pour tous les dénombrements qu'elles permettent dans la phase d'analyse.
Les balises de locuteur se présentent sous la forme suivante: « [>NOM]: ». Sonal recherche les séquences « [> » puis «]: » et considère ce qu'il y a entre comme le nom des locuteurs.

Les balises qui se terminent par un point d'interrogation sont considérées comme des questions (par exemple : « [>Question ?]: »)
Pour que cela fonctionne bien, **il ne faut pas qu'il y ait d'espace dans le nom** des locuteurs et il faut que la balise se termine bien par «]: » et nom par «] : ».

- Les points de synchronisation (normaux ou dérivants)

Les points de synchronisation sont extrêmement utiles. Il ne faut pas hésiter à en mettre souvent en phase de retranscription. Ils apparaissent dans le texte sous la forme d'un bouton et sauf si vous en placez tous les trois mots, cela ne compliquera pas votre utilisation du logiciel, au contraire.

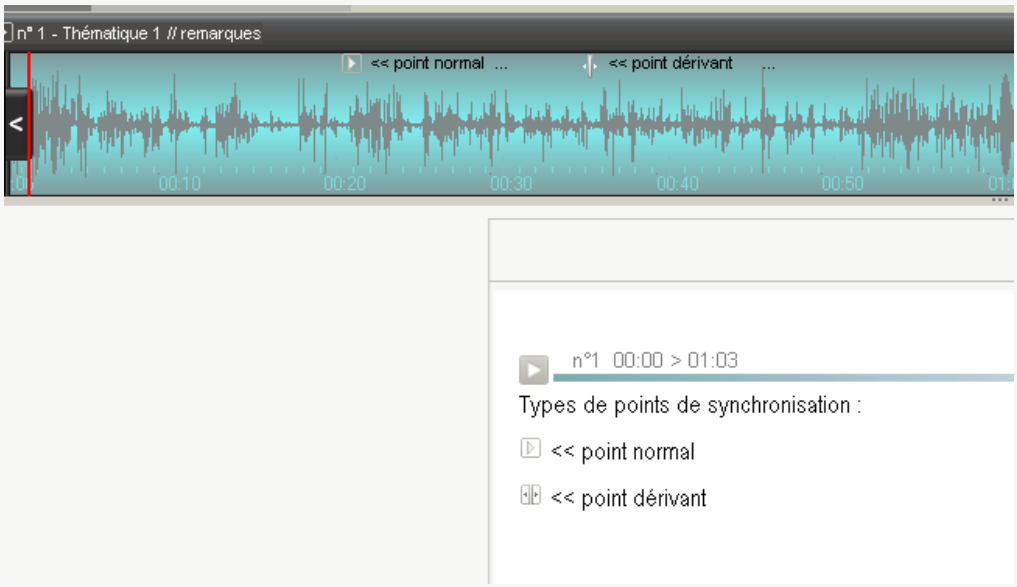
Concrètement, il s'agit de balises entre crochets placées dans le texte contenant les **coordonnées temporelles (en secondes)** des passages : [XX,XX]. Par exemple « [95,0] » signifie que le passage qui suit ce point est situé à 01m35s.

Ils sont pratiques pour repérer les phrases intéressantes que l'on veut pouvoir réécouter ou voir ressortir. Mais au-delà, les points de synchronisation permettent surtout d'offrir une plus grande souplesse dans l'utilisation du logiciel. Ils permettent notamment de redécouper les extraits (cf. plus bas) et améliorent le calcul de position. Il y a donc un grand intérêt à en placer très régulièrement dans le texte.

Normalement, si vous retranscrivez dans Sonal, des points seront ajoutés un peu partout au fil de votre travail. Mais vous pouvez en placer manuellement.

Il existe un type particulier de points de synchronisation : les points dérivants. Sonal les interprète exactement comme des points normaux mais ils ont la propriété de « dériver » selon les corrections apportées aux points qui les entourent. Ils sont ajoutés avant chaque changement de locuteur lors de la « resynchronisation » de bandes sonores et de leur retranscription. Leur position est estimée sur la base des coordonnées connues environnantes. Ils devront ensuite être corrigés, avec ceci de pratique que les corrections apportées à un point se « propagent » aux suivants.

Ces points se caractérisent par la présence d'un point d'interrogation à la fin. Ex : [248,0?] signifie qu'au dernier calcul, le passage considéré était estimé à 04m08s. Pour signaler cette incertitude, ils sont représentés d'une manière spécifique.



Les points sont dits dérivants car ils s'adaptent à toute modification de la trame temporelle environnante. Concrètement, si vous déplacez un point vers la gauche ou la droite, tous les points avant et après se décalent dans le même sens. De cette manière, **on décale d'un même mouvement plusieurs points** et, logiquement, l'estimation se fait de plus en plus précise à mesure que l'on corrige les points. Dans certains contextes (voir plus bas), ce procédé est vraiment très efficace.

Sélection d'un point de synchronisation. Décalage vers la droite

Déplacement de tous les points dérivants environnants :

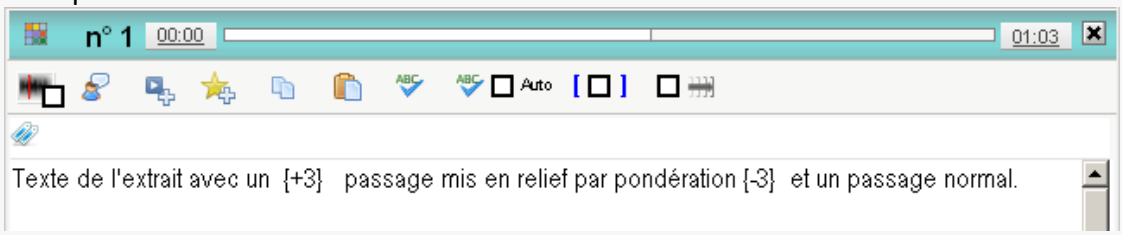
- Les formats de pondération

Sonal vous permet de repérer dans le texte certains passages que *vous* jugez intéressants ou illustratifs, et de les faire ressortir par une mise en forme particulière allant d'une taille de caractère de +1 à +5 :

L'intérêt de cette pondération qualitative est qu'elle pourra être utilisée pour filtrer les extraits et leur contenu textuel : en phase d'analyse thématique vous pourrez ne voir et n'exporter que les passages pondérés à un certain niveau.

Il s'agit là encore de simples balises laissées dans le texte : le poids du passage est signalé au début par {+x} et à la fin par {-x}, x pouvant prendre les valeurs de 1 à 5.

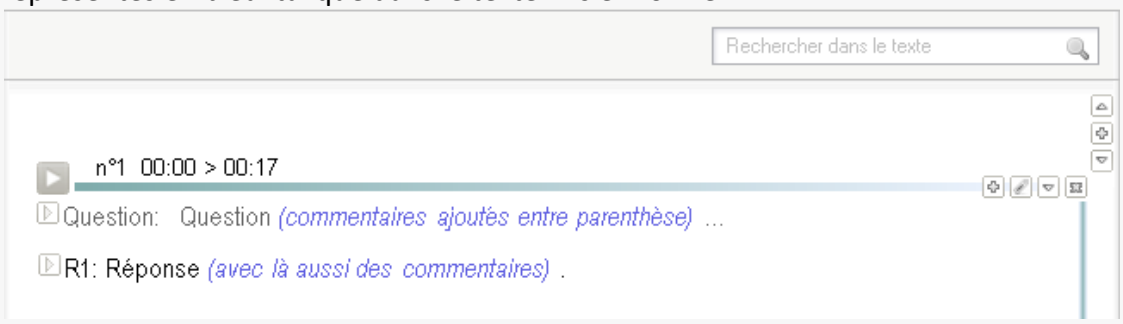
Exemple :



Donne une fois interprété :

- Les commentaires

Enfin, il est utile de savoir que **Sonal considère comme des commentaires tout ce qui apparaît entre parenthèses dans les retranscriptions.** Les commentaires sont représentés en bleu italique dans le texte mis en forme :



Les commentaires peuvent être affichés ou masqués lors de la lecture et ils sont exclus des dénombrements lors des analyses lexicométriques.

Construire un corpus

3 Se lancer !

3.1 Première étape : créer un répertoire de projet[3]

Pour commencer votre travail et construire votre corpus, avant même le lancement du logiciel, vous devez **créer ou choisir un répertoire qui accueillera votre projet**.

- Où installer ce répertoire de projet ?

Pour créer votre répertoire de projet, **choisissez un emplacement où vous disposerez de suffisamment de place et où l'accès est rapide** ; évitez de travailler directement sur un support externe de type USB, cela ralentit un peu le logiciel et comporte plus de risques d'accident (retrait involontaire de la clé, panne, etc). Utilisez plutôt ces supports pour conserver des sauvegardes (cf. plus haut).
Par ailleurs, assurez-vous que vous disposez de droits en écriture dans le répertoire en question. Sur les ordinateurs professionnels, il n'est pas rare que certaines parties du disque soient protégées. En pareil cas, Sonal risque de ne pas fonctionner correctement.

Placez tous les enregistrements audio/vidéo et tous les documents que vous voulez traiter dans ce dossier.

Avant de commencer à traiter vos entretiens, il va falloir créer le projet qui va les accueillir, et qui contiendra toutes les informations transversales sur vos entretiens. C'est ce qu'on appelle un corpus.

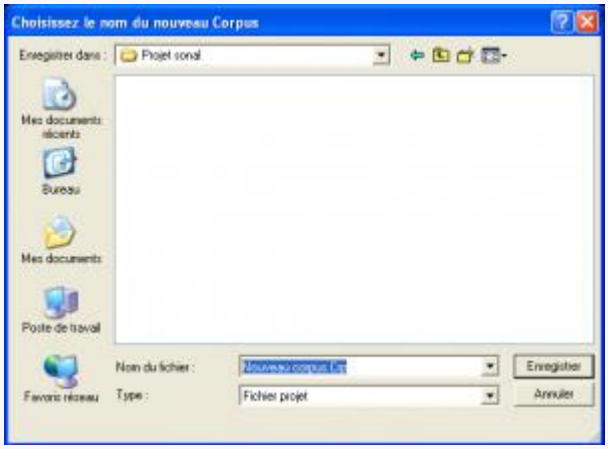
3.2 Créer un nouveau corpus

D'une manière générale, un corpus est une addition de documents (media ou texte) relatifs à un même sujet, que l'on agrège à des fins d'archivage ou d'analyse. Dans l'usage le plus courant de Sonal, c'est un ensemble d'entretiens enregistrés dans le cadre d'une enquête.

Pour créer un nouveau corpus, il suffit de lancer Sonal. Sur la page d'accueil, sélectionnez « Nouveau corpus» (cf. ci-dessous)

[3] La version en ligne de ce tutoriel contient des vidéos qui pourront peut-être vous aider.

La boite de dialogue qui s'ouvre alors correspond à un « enregistrer sous » :

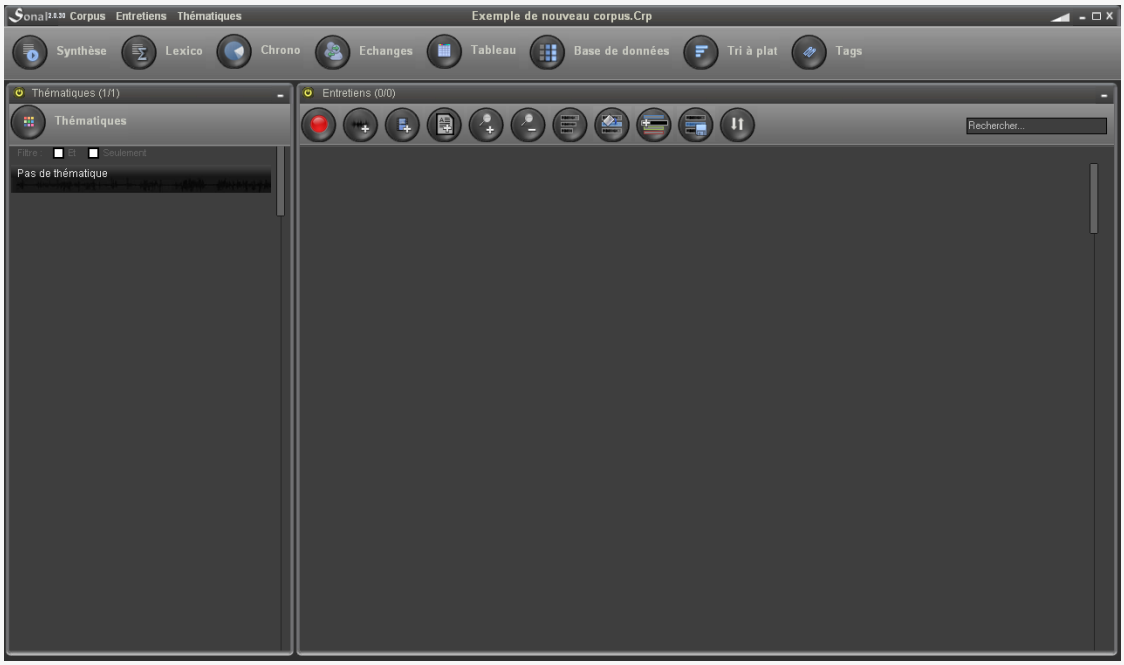


Sonal vous demande comment s'appellera et où sera positionné le fichier corpus que vous voulez créer. **Il faut l'enregistrer dans le répertoire de projet, celui qui contient déjà vos entretiens**.

> *Dans la mesure où la boite de dialogue recherche spécifiquement des fichiers dont l'extension est *.crp, vous devez normalement ne voir aucun fichier dans votre dossier de projet, même s'il contient vos bandes. Vous devez donc choisir un nom et cliquer sur « enregistrer ».*
> *Si vous voyez des fichiers corpus dans le dossier, sachez que les sélectionner à cette étape revient à demander à Sonal de créer un nouveau projet en les écrasant. Logiquement, Sonal refusera de faire une telle chose.*

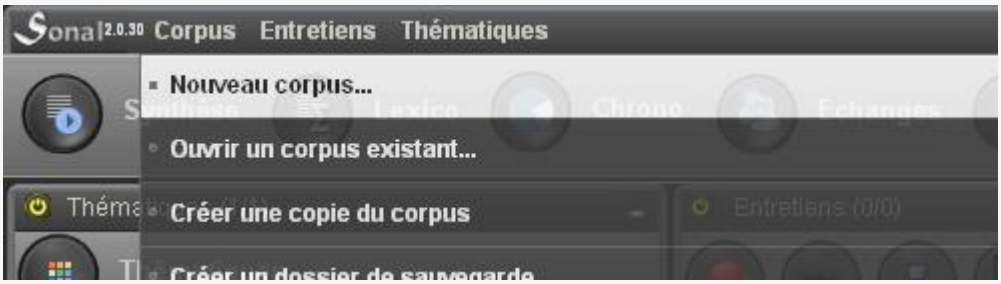
Choisissez un emplacement et un nom puis validez.

Ceci fait, vous avez un projet vierge. La page principale a cette apparence :



Le projet de ne contient ni entretien, ni variable, ni thématique, à l'exception de la thématique vierge (sans couleur) qui répond au nom de « Pas de thématique » et qui figure dans le panneau gauche de l'écran, sous le bouton « thématiques ».

> *Si vous ne voyez pas cette thématique, c'est qu'il n'y a pas de projet ouvert. Aucune fonctionnalité ne répondra, tant que vous n'aurez pas ouvert ou créé un corpus. Pour ce faire, allez dans le menu corpus :*



Il va donc falloir commencer à construire votre corpus.

- Les étapes de la construction du corpus

Construire un corpus Sonal veut dire faire plusieurs choses. Un corpus complet contiendra :

1) des documents sonores ou textuels

2) un dictionnaire des variables (base de données)

3) une grille thématique.

4) des fichiers d'encodage associés à chaque document (Rtr)

> *Si l'ajout des documents à encoder va de soi, la construction du dictionnaire des variables peut intervenir à tout moment, soit avant soit après l'encodage. La grille thématique gagne quant à elle à être réalisée aussi tôt que possible, et au plus tard en cours de retranscription, car il est précieux de commencer l'encodage thématique avec la retranscription. C'est l'un des avantages de Sonal, il serait dommage de s'en priver.*

Nous allons commencer par apprendre à ajouter des bandes ou des documents texte à notre corpus, avant de voir comment construire la grille thématique. Nous verrons ensuite comment construire et renseigner la base de données, avant de rentrer dans le travail d'encodage proprement dit, avec la retranscription et le découpage sonore.

4 Ajouter des documents à votre corpus

- Types de fichiers compatibles

Sonal permet d'encoder 3 types de documents :

- **Des documents sonores.** Sonal travaille préférentiellement avec des documents **WAV (PCM)** encodés au format 16 bits mono, mais supporte également les formats compressés de type *.**MP3, *.WMA, *.OGG, *.AMR, *.M4A**. La différence est que Sonal ne peut dessiner que les fichiers WAV :

  Cette différence est seulement d'ordre graphique mais, pour positionner le lecteur et repérer les changements de parole, il est préférable de voir le spectre sonore. En phase de retranscription/codage, le format WAV est donc préférable, même si les fichiers non compressés sont évidemment plus lourds. Une manière astucieuse de travailler peut être d'utiliser le WAV le temps de l'encodage, et de lui substituer un format compressé par la suite. Pour quelques conseils sur la conversion des fichiers sons, reportez-vous à ce tutoriel.

- **Des vidéos.** Sonal supporte les formats ***.AVI et *.MPG**, sous réserve d'installer les codecs nécessaires pour que Windows Media Player soit capable de les lire.

- **Des fichiers texte.** Sonal peut soit importer des documents depuis Word ***.DOC, *.DOCX**, soit depuis des fichiers au format ***.TXT (ANSI)**. Dans les deux cas, ceux-ci seront convertis en texte simple (ANSI).

- Possibilités d'encodage

Ces trois types de documents font l'objet d'un même type d'encodage. Au-delà des valeurs renseignées dans la base de données, leur contenu sera découpé en parties successives, contigües ou non, mais de préférence non superposées.

> *Pour les sons ou les vidéos, les documents sont découpés sur la base des coordonnées chronométriques (en secondes), pour les documents texte, c'est la position du mot de départ rapportée au total des mots qui sert de référence (en millièmes).*

Les extraits découpés se verront associer une ou plusieurs couleurs thématiques et du texte libre, soit des retranscriptions, soit toutes sortes d'annotations (commentaires, données musicales etc.).

4.1 Enregistrer un document sonore

Le fait d'utiliser Sonal comme enregistreur présente plusieurs avantages :

- les documents sonores sont directement intégrés au corpus courant;
- il est possible de délimiter les extraits en cours d'enregistrement, par simple utilisation de la touche entrée;
- il est possible d'ajouter des notes aux extraits découpés;
- il est possible d'affecter des thématiques aux extraits en cours d'enregistrement.

Cette méthode est particulièrement adaptée à l'enregistrement de séminaires, de conférences, voire de cours.

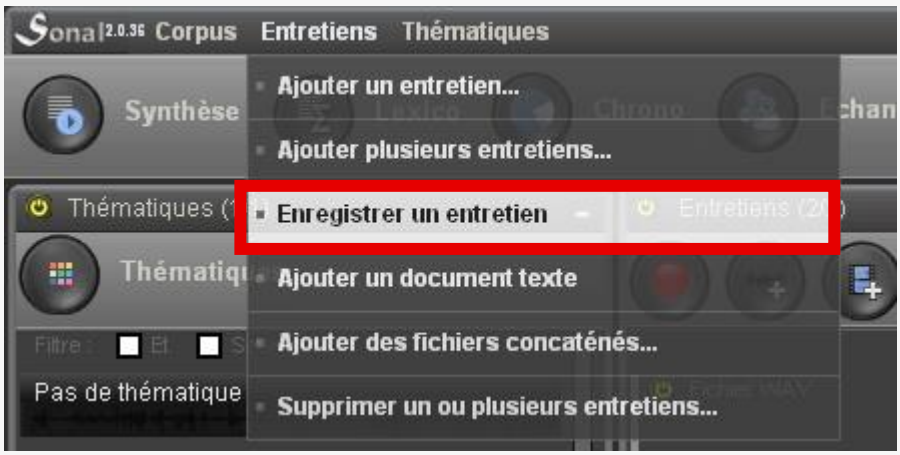
Attention, cette fonctionnalité repose sur un dialogue avec la carte son qui n'est pas également possible sur toutes les machines (il y a eu longtemps un problème avec certaines versions de Windows, qui doit normalement être corrigé dorénavant). Si Sonal n'y arrive pas sur votre PC, vous le saurez rapidement... ☺

Voici comment vous y prendre pour enregistrer un entretien dans Sonal.

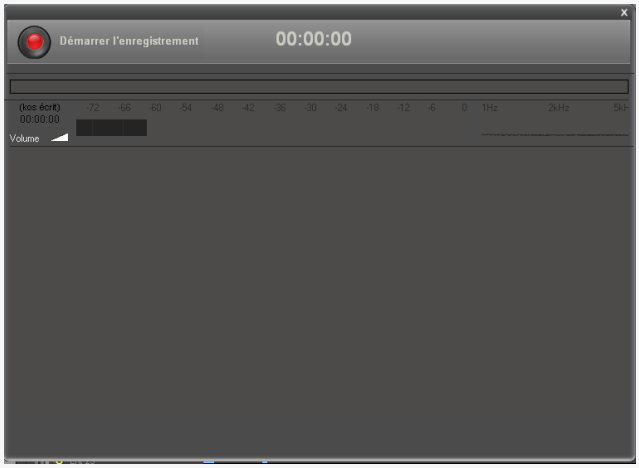
> *Avant toute chose, si vous ne l'avez pas déjà fait, **vous devez créer ou ouvrir un corpus**. Ce corpus définira notamment l'endroit où sera enregistré le fichier que vous allez créer (il sera positionné dans le même dossier que le fichier *.Crp). Il vous fournira également les thématiques que vous pourrez affecter aux extraits isolés en cours d'enregistrement (cf. ci-dessous).*

L'enregistrement d'un nouvel entretien se lance avec ce bouton :

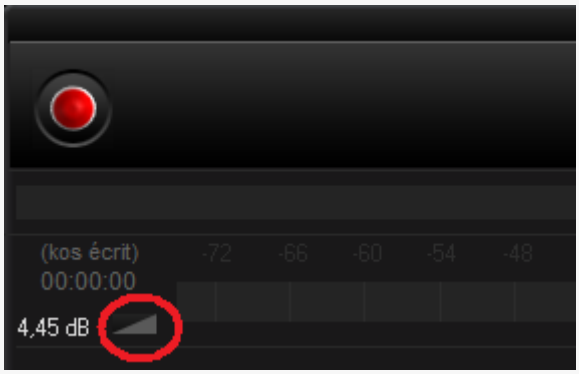
ou depuis le menu Entretiens:



Vous arrivez sur la page d'enregistrement :



La première chose à faire est de **régler le volume sonore.** Lorsque le curseur indique un volume adapté (ni trop faible, ni saturé), vous pouvez démarrer l'enregistrement.


(kos écrit)
00:00:00
-72
-66
-60
-54
-48
4,45 dB

La fenêtre de définition des caractéristiques de l'enregistrement apparaît :

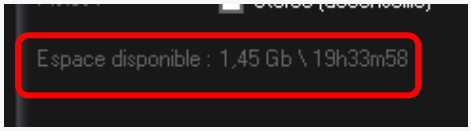
Pour pouvoir lancer l'enregistrement, vous devez impérativement **nommer le fichier** que vous allez enregistrer.

Cette page vous permet également de définir les caractéristiques du fichier :

- la fréquence (11025Hz par défaut)

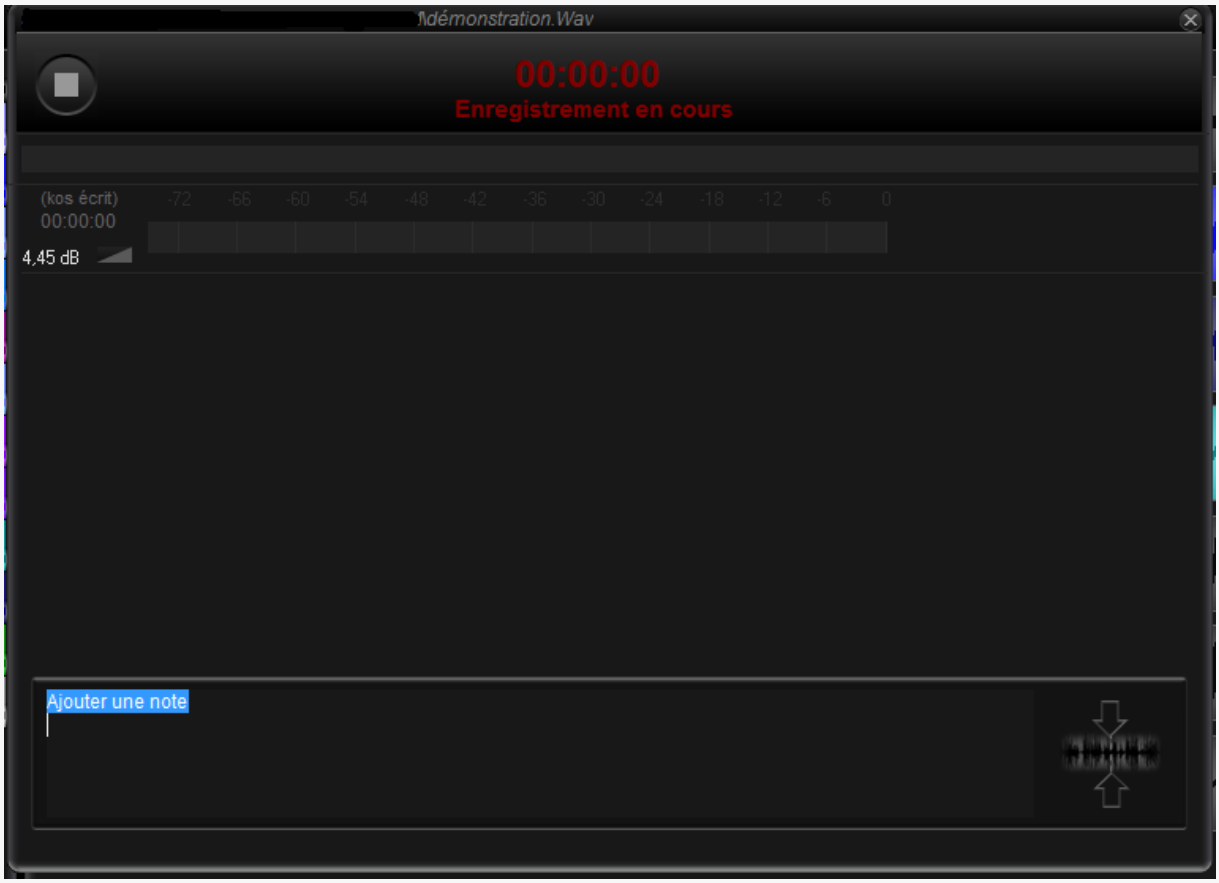
- le nombre de canaux (Mono par défaut, la stéréo est possible mais déconseillée dans le cas d'enregistrement d'entretiens)

L'interface vous informe sur l'espace restant sur le disque ou le périphérique où vous allez enregistrer votre fichier et vous précise combien de temps de bande peut être enregistré (en fonction du format utilisé).



Après avoir défini le nom du fichier et choisi son format, vous pouvez lancer l'enregistrement.

L'interface prend alors cette apparence :

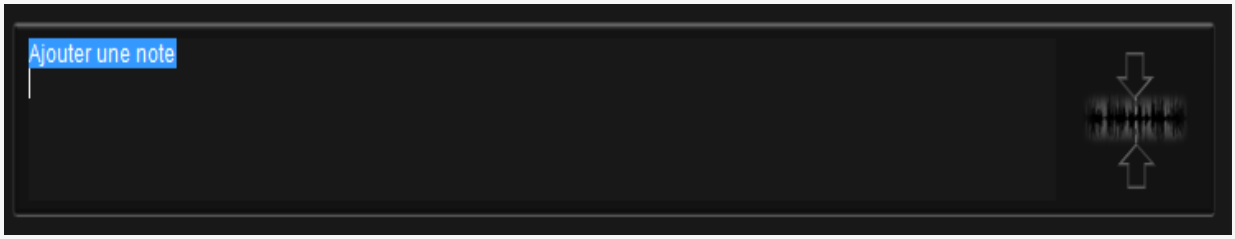


Vous pouvez alors laisser "tourner la bande" jusqu'à la fin de l'enregistrement, que vous signifierez simplement en cliquant sur le bouton stop.

Une fois l'enregistrement terminé, l'entretien est automatiquement ajouté au corpus courant et chargé.

Si vous souhaitez gagner du temps dans vos retranscriptions, vous pouvez placer des marqueurs en cours d'enregistrement, c'est-à-dire prédéfinir les extraits et éventuellement leur affecter une thématique.

Pour ce faire, vous pouvez cliquer sur le bouton ci-dessous, ou simplement **taper entrée**, en notant que le curseur doit être dans la zone de texte vouée à la saisie des notes pour que la touche entrée fonctionne :



Ce faisant, vous définissez la fin de l'extrait, auquel sera ajouté le texte contenu dans la zone de texte. Les différents extraits découpés apparaissent par la suite par ordre chronologique inversé.

Si vous avez déjà défini des thématiques, vous pouvez directement les affecter aux extraits que vous avez découpés en cliquant sur les petites flèches en bout de ligne:

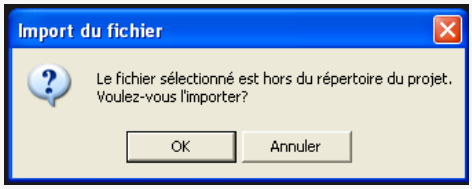
4.2 Ajouter un document sonore

On ajoute un document sonore soit directement depuis le bouton ci-dessous :

Ou à travers le menu ajouter un entretien (ou un enregistrement)

Apparaît alors une boite de dialogue de type « Ouvrir » dans laquelle il vous faudra choisir le fichier son que vous voulez importer.

Si le fichier est hors du répertoire de projet, qui contient votre fichier corpus, Sonal vous demandera de l'importer :



> *Le processus prendra plus ou moins de temps en fonction de la taille du fichier mais Sonal ne vous renseignera pas sur la progression de la copie. Vous aurez seulement un message de confirmation à la fin de la copie.*

Votre fichier est ensuite ajouté au corpus. S'il est au format WAV, Sonal va d'abord prendre le temps de le dessiner intégralement, ce qui peut prendre quelques secondes. S'il est dans un format compressé, Sonal se contentera de dessiner un spectre sonore fictif et renseignera le format du fichier au début de l'image (voir ci-dessous).

4.3 Ajouter une vidéo

Pour ajouter une vidéo (format *.AVI ou *.MPG) Cliquez sur le bouton ci-dessous :

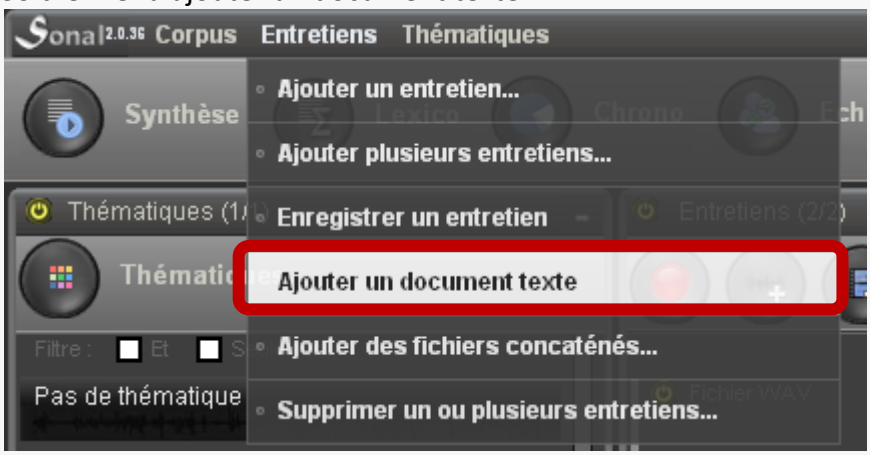
Vous pouvez aussi suivre la même méthode que pour ajouter un fichier son, en sélectionnant « Fichiers vidéo » dans le filtrage des types de fichier.

> *Le fichier vidéo sera représenté comme un fichier audio compressé, et Sonal ne pourra pas le dessiner. Pour pouvoir travailler les vidéos dans de meilleures conditions, vous pouvez d'abord ajouter l'extraction sonore de la vidéo au format WAV, et de placer* ***dans le même répertoire*** *la version vidéo, avec* ***strictement le même nom****, hormis l'extension (qui doit être *.Avi ou *.Mpg.) Sonal repérera cette vidéo et l'utilisera dans le lecteur, mais il dessinera la version WAV qui lui est associée.*

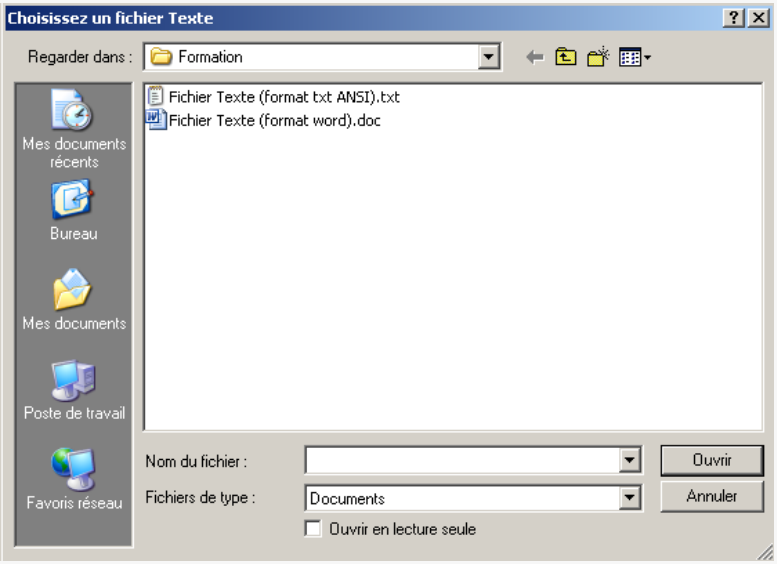
4.4 Ajouter un fichier texte

Pour importer des fichiers textuels dans Sonal (retranscriptions seuls, documents divers), il faut utiliser soit le bouton ci-dessous :

Soit le menu ajouter un document texte :



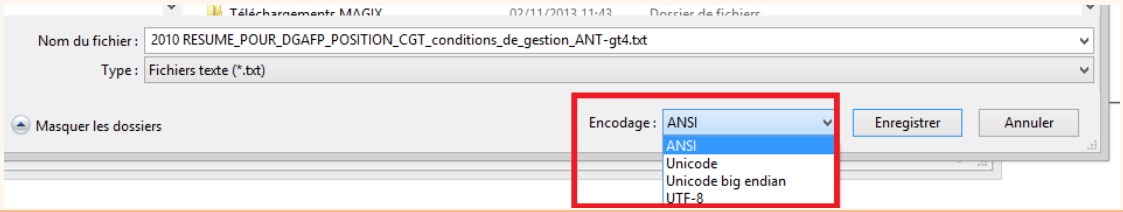
La boite de dialogue qui apparaît recherche des fichiers Word (*.Doc, *.Docx) ou des fichiers Texte (*.TXT) :



Ceux-ci peuvent être placés n'importe où sur l'ordinateur, Sonal se contentera d'en recopier leur contenu.

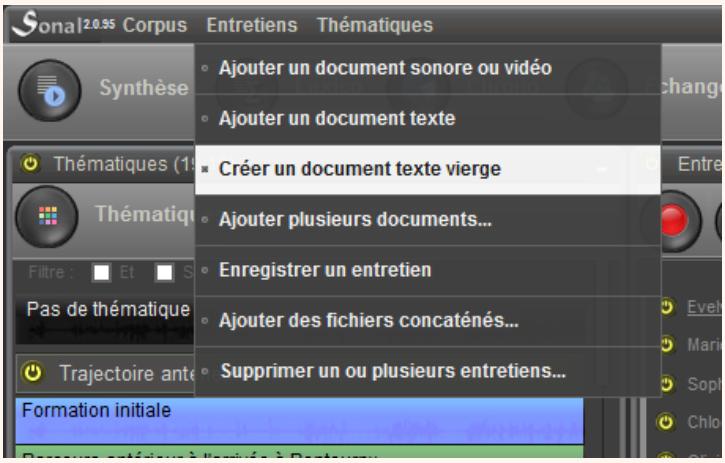
> *Attention, les fichiers Word ne seront correctement importés **que si Word est installé sur votre machine**. Sonal dialogue en effet avec le logiciel lui-même. Si vous avez un traitement de texte libre, utilisez-le pour exporter le contenu textuel des fichiers word vers un fichier *.Txt (créé par exemple avec le Bloc Note). Le résultat est strictement le même que si vous l'aviez importé depuis word.*

Attention, lors de l'enregistrement de votre fichier TXT, pensez, dans les options, à spécifier un codage au format ANSI :

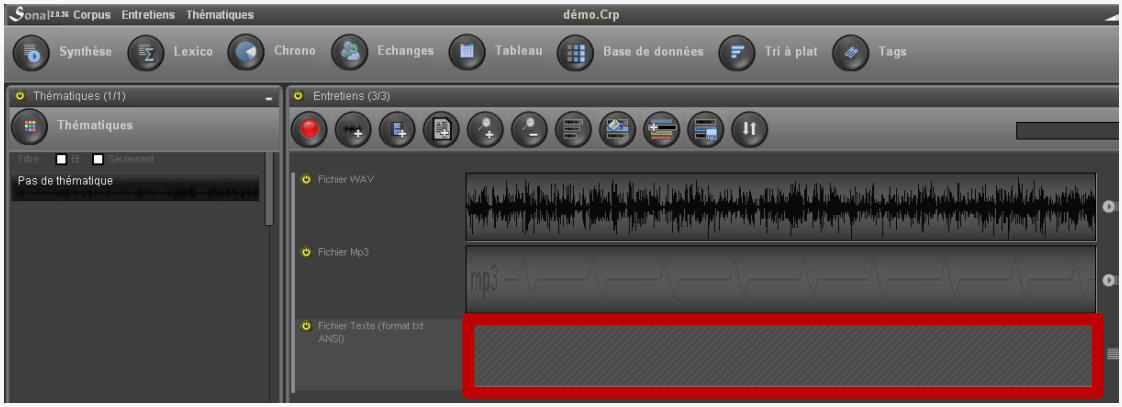


Autre méthode, pour éviter d'avoir à créer un fichier txt (depuis la version 2.0.80) :

Depuis la page de corpus, vous pouvez **créer un document texte vierge** (Menu 'Entretiens' -> 'créer un document texte vierge', l'ouvrir et y coller tout le texte du document que vous souhaitiez importer.



La représentation des fichiers texte suite à leur importation est un peu différente de celle des sons ou des vidéos. Sonal crée d'office un seul très long extrait non thématisé qu'il faudra découper par la suite (voir ici)

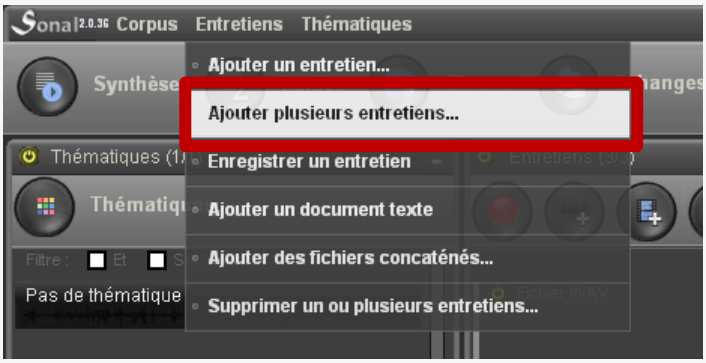


4.5 Ajouter un fichier Rtr

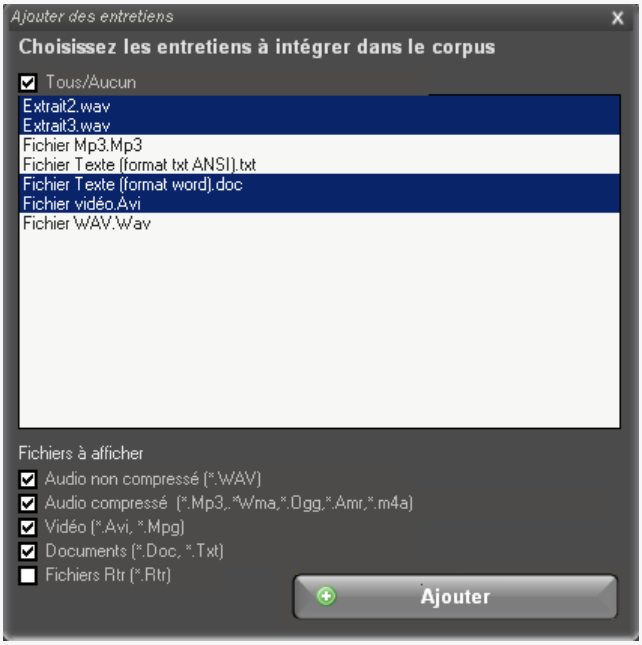
Lorsque vous ajoutez un document texte dans Sonal, celui-ci ajoute directement à la liste des entretiens (dans le fichier CRP) un fichier *.Rtr (normalement, ce sont les noms des fichiers audio/vidéo qui sont renseignés, à partir desquels Sonal cherche les fichiers Rtr). En chargeant le corpus, Sonal sait donc qu'il n'y a pas de fichier audio/vidéo à chercher pour la piste en question, et ouvre directement le Rtr.
Jusqu'à la version 1.8, cela rendait difficile l'importation dans un corpus de fichiers textuels déjà encodés. Faute d'avoir une bande à cibler pour l'importation, on ne pouvait pas importer directement de Rtr. On peut dorénavant le faire via l'importation de plusieurs documents simultanément (voir ci-dessous). Une case à cocher dans l'ajout simultané de plusieurs documents vous permettra d'ajouter des Rtr. Faites seulement attention de ne pas ajouter les Rtr de fichiers déjà présents dans le corpus et mémorisés par le nom de leur enregistrement.

4.6 Ajouter plusieurs documents simultanément

Pour gagner du temps, il est possible d'ajouter en une seule fois un grand nombre de documents, quel que soit leur format. Pour cela, utilisez le menu ci-dessous :



La fenêtre qui apparaît alors **inspecte le contenu du répertoire de projet**, et vous propose d'importer tous les fichiers présents répondant au filtrage voulu.



Sélectionnez les fichiers voulus et cliquez sur ajouter. Tous les fichiers seront ajoutés successivement.

> *Si vous cliquez sur la case « Tous/aucun », Sonal sélectionnera tous les fichiers qui ne sont pas encore présents dans le projet.*

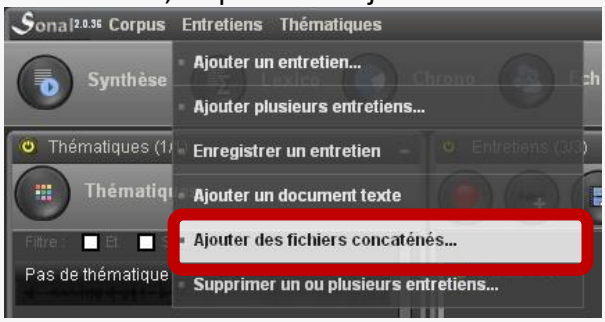
4.7 Concaténer des fichiers son

Il est assez fréquent que l'on doive, au cours d'un entretien, arrêter temporairement l'enregistrement. En pareil cas, on obtient plusieurs fichiers pour un seul entretien, ce qui est un peu dommage. La solution est de **« recoller » les bandes isolées les unes au bout des autres pour reconstituer un long enregistrement**. C'est cette tâche que l'on appelle la « concaténation » des fichiers.

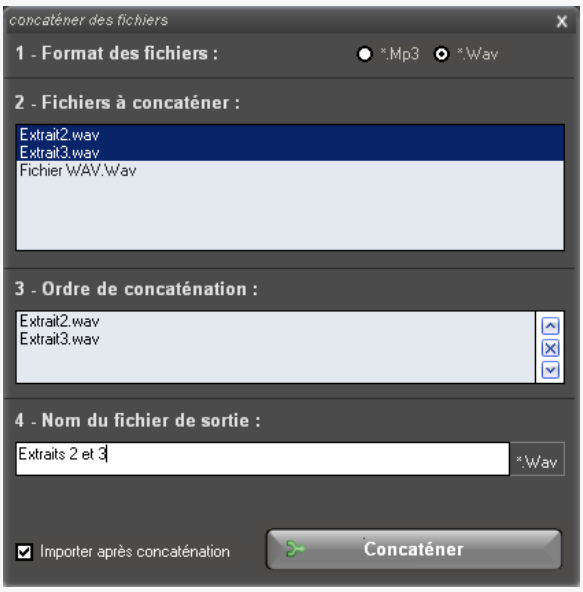
Sonal vous permettra de concaténer vos fichiers à plusieurs conditions. Pour utiliser les fonctionnalités proposées par Sonal, il faudra :

- Que vos fichiers soient au format WAV ou MP3
- Qu'ils soient de même format (s'ils proviennent d'un même enregistreur cela devrait être le cas)
- Qu'ils soient stockés dans le même dossier que le fichier corpus

Si ces conditions sont remplies, voici comment vous y prendre. Dans le menu « entretiens » ou « documents », cliquez sur « Ajouter des fichiers concaténés » :



L'interface a cette apparence :



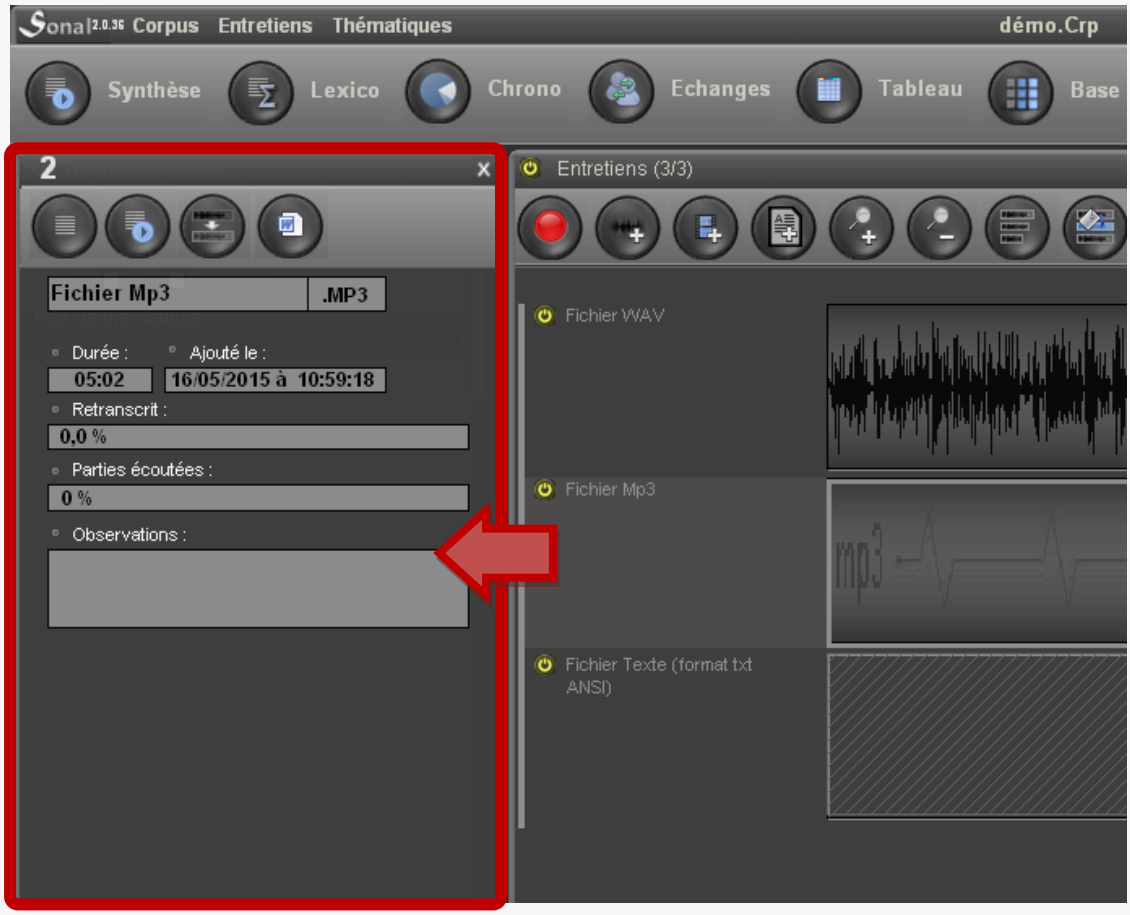
Il faut en suivre les étapes. A savoir
1) commencer par choisir un format de fichier (MP3 ou WAV)
2) cliquer dans la liste des fichiers (tous les fichiers au format donné trouvés dans le répertoire du projet).
3) définir l'ordre de concaténation, voire retirer certains fichiers à l'aide des boutons situés à droite de la liste
4) à nommer le fichier de sortie qui réunira les fichiers choisis dans l'ordre choisi.

Après dix secondes laissées au système pour réaliser la concaténation, votre nouveau fichier va être importé.

5 Gérer les documents de votre corpus

Cette partie présente les différentes manipulations qu'il est possible de faire sur les documents d'un corpus une fois ceux-ci ajoutés.

Un certain nombre de fonctionnalités présentées ci-après sont proposées dans un **panneau latéral** qui **apparaît lorsque l'on clique sur le <u>nom de l'entretien</u> dans le corpus** :



Apparaissent alors plus d'informations sur l'entretien et des outils pour le modifier.

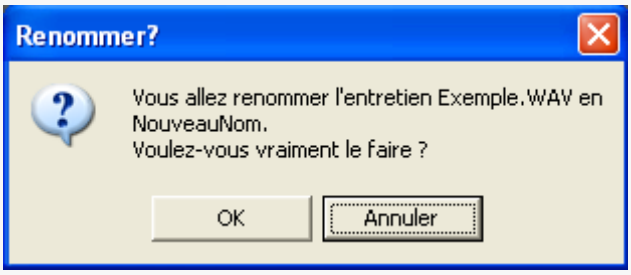
5.1 Renommer un document

On a souvent besoin de changer le nom d'un entretien, notamment pour protéger l'anonymat des répondants. Dans la mesure où Sonal crée beaucoup de fichiers portant le même nom et que c'est ce nom commun qui permet au logiciel de les retrouver, il faut être très prudent lorsque l'on renomme les fichiers. Pour éviter les mauvaises manipulations, Sonal vous propose une procédure simple.

Dans la fenêtre d'informations sur l'entretien (que l'on obtient en cliquant sur le nom de l'entretien dans la page de corpus), la zone dévolue au rappel du nom du fichier est une zone de texte :

Pour changer le nom du document et de tous les fichiers créés à partir de lui, il suffit de **taper un nouveau nom et de valider avec entrée**.

Sonal vous demande une confirmation :



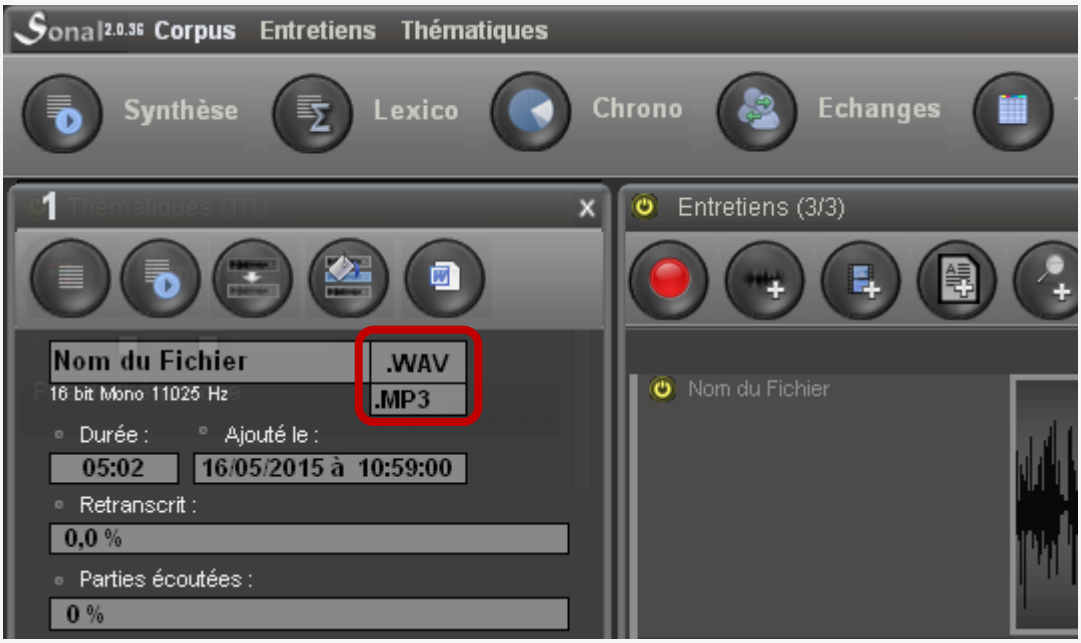
Après acceptation, l'entretien, ainsi que tous les fichiers satellites sont renommés.

5.2 Changer de format de document (allégez vos corpus)

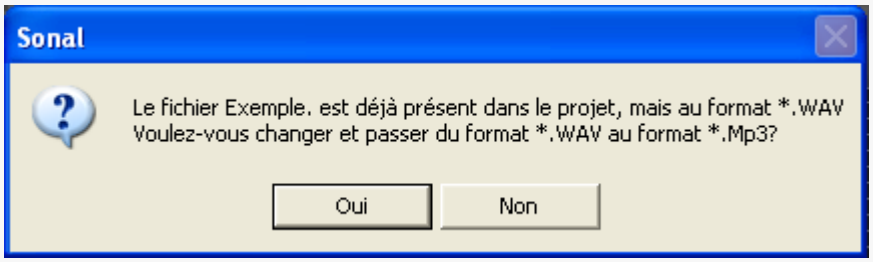
S'il est préférable de travailler avec des fichiers WAV en phase d'encodage, ceux-ci deviennent inutilement encombrants en phase d'analyse. Si vous tenez à conserver vos bandes dans le processus d'analyse, il peut être intéressant de substituer aux fichiers WAV des fichiers plus légers (Mp3, Wma) etc.

> *Vous pouvez également opter pour un travail sur « corpus muet », qui sera beaucoup plus léger.*

La méthode la plus simple pour y arriver est de **placer dans le répertoire du projet une version compressée du ou des fichiers que vous voulez remplacer. Ils doivent porter strictement le même nom.**
En haut du panneau latéral est renseigné le nom du document. Une case située immédiatement à droite du nom renseigne son extension. Lorsque l'on clique dessus, les autres extensions disponibles dans le corpus sont répertoriées. Dans l'exemple ci-dessous, il suffit de cliquer sur le Mp3 pour que le document associé à l'entretien en question soit la version mp3 et non plus WAV du fichier. Cette dernière pourra être supprimée ou en tout cas retirée du dossier de projet sans que cela change quoi que ce soit pour Sonal.



Une autre méthode peut être de demander l'ajout d'un document sonore et de choisir la version compressée du fichier que vous voulez remplacer. Sonal comprendra que vous voulez effectuer le remplacement et vous demandera une confirmation :

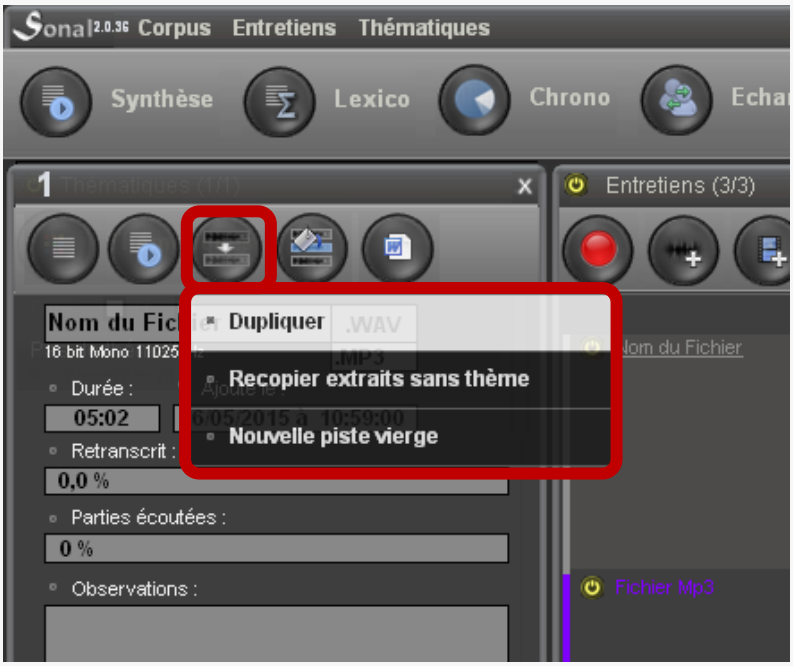


5.4 Dupliquer une piste

Il est possible d'avoir plusieurs pistes d'encodage pour un même entretien. Concrètement, plusieurs fichiers Rtr se réfèreront à une seule et même bande. Cela peut vous permettre d'encoder différents aspects d'un même entretien, de confronter les encodages de plusieurs enquêteurs etc. Les usages potentiels sont très nombreux.

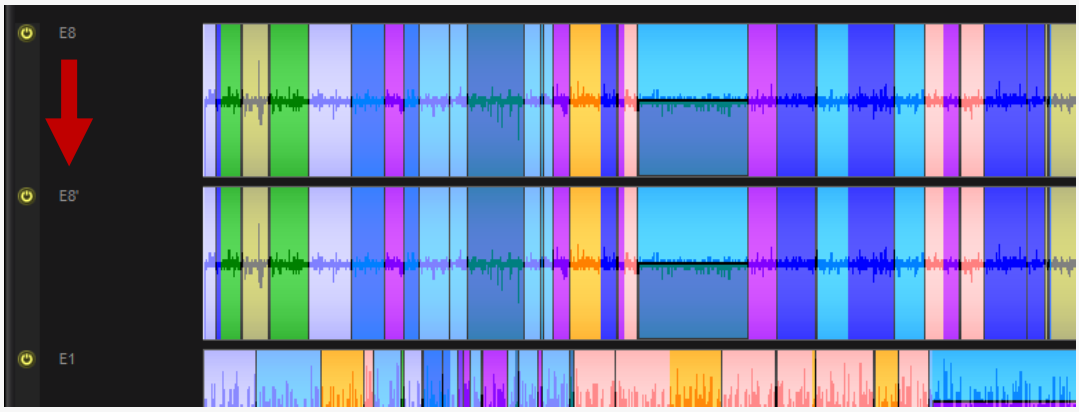
> *Le fonctionnement est simple : lorsque vous créez une nouvelle piste pour un enregistrement donné, celle-ci porte le même nom que la piste d'origine mais avec un prime à la fin.*
>
> *Si vous dupliquez une piste " exemple ", la seconde s'appellera " exemple' "; si vous ajoutez une troisième, elle se nommera " exemple'' " et ainsi de suite. Lorsqu'il lira ces pistes, sonal retirera automatiquement les ' à la fin du nom et chargera le fichier son d'origine.* Cela implique qu'il faut faire attention lorsque l'on souhaite renommer les pistes : changer le nom de la piste d'origine ou des pistes ajoutées rendra impossible la reconnaissance du fichier d'origine.

Pour dupliquer une piste, il faut d'abord la sélectionner dans le corpus (en cliquant sur son nom). La fenêtre d'informations apparaît alors sur le côté gauche de la page. Dans cette fenêtre il faut cliquer sur le bouton ci-dessous :

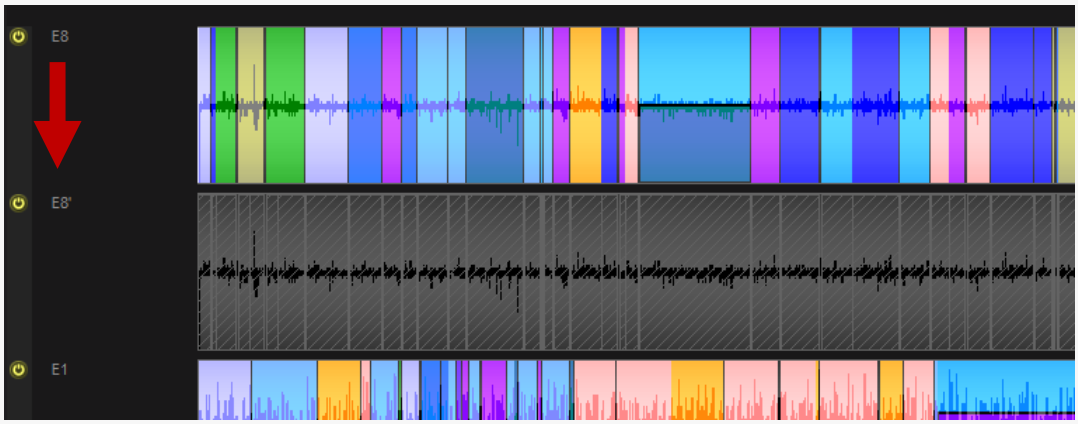


Vous avez alors trois possibilités :

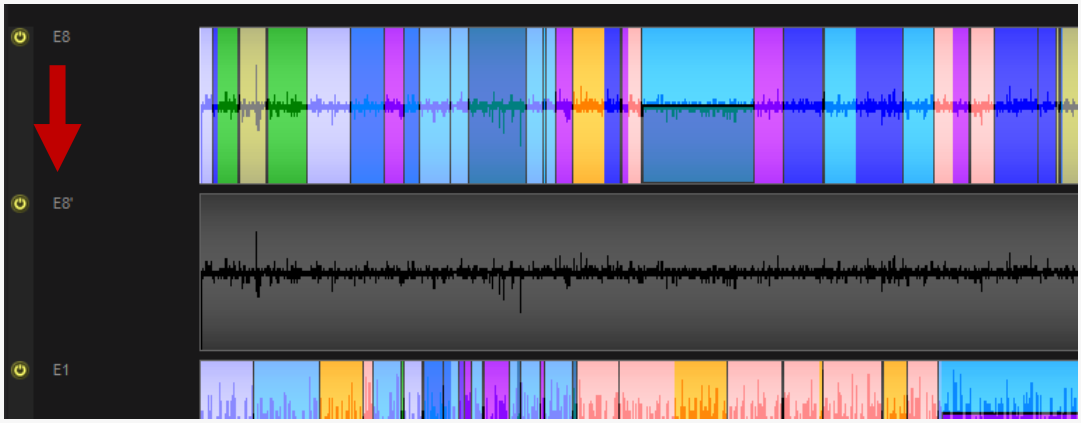
Dupliquer la piste reviendra à la recopier intégralement (extraits et thématisation). Vous la clonez en quelque sorte :



- **Recopier extraits sans thème** revient à dupliquer la structure des extraits mais à les laisser vierges de toute thématique:



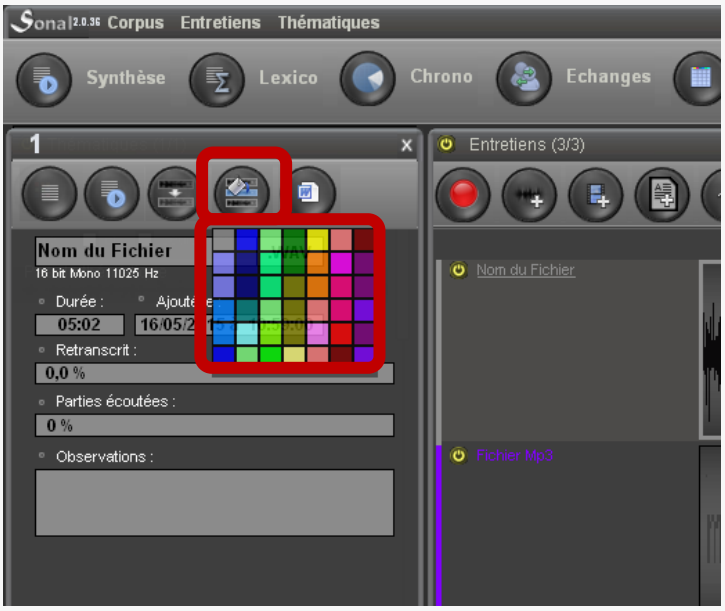
- **Nouvelle piste vierge** vous permet comme son nom l'indique d'avoir une nouvelle piste pour encoder librement le document en question.



5.5 Colorier les documents

Il est possible d'affecter un code couleur à vos entretiens, soit individuellement, soit collectivement. Ces codes couleurs peuvent vous permettre de retrouver plus facilement vos entretiens.

Pour colorier un entretien individuellement, allez dans le panneau latéral :

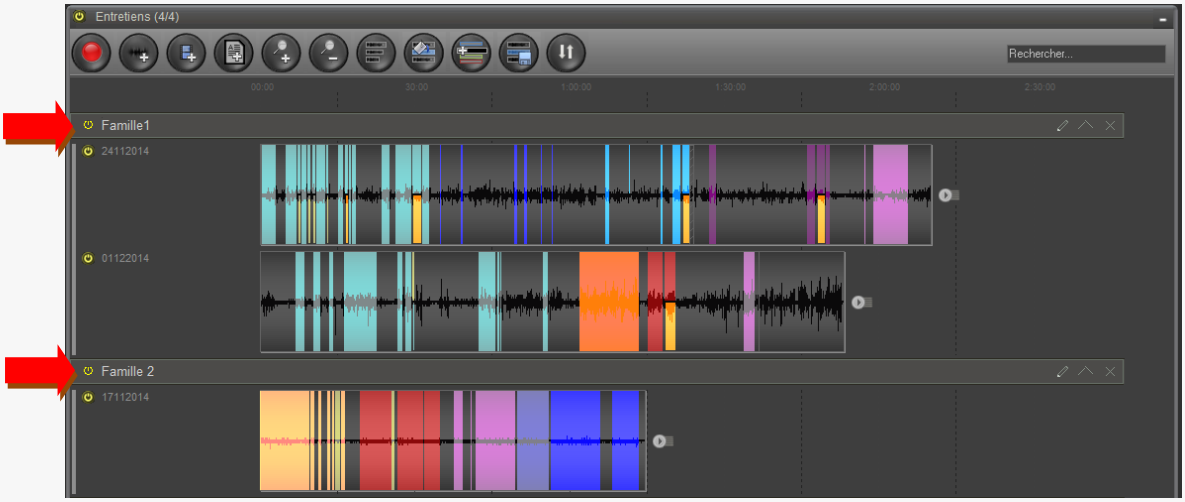


Pour colorier plusieurs entretiens en une seule fois, cliquez sur le bouton ci-dessous et choisissez une couleur.

Tous les entretiens ***actifs*** seront coloriés de la couleur choisie.

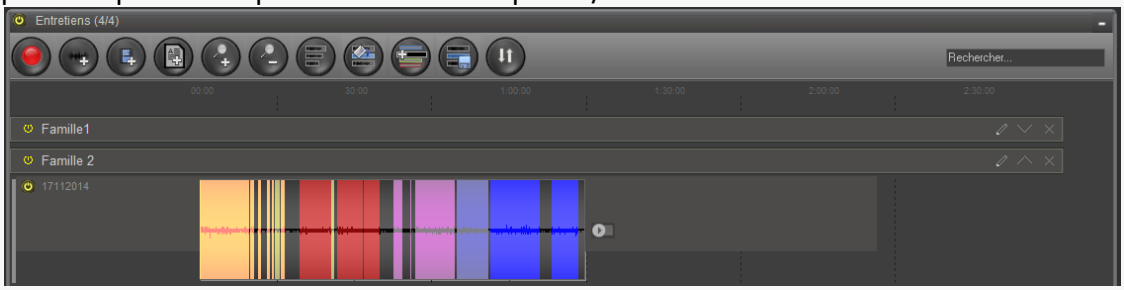
5.6 Ajouter des séparateurs (créez des « familles »)

Sonal 2 vous permet d'ajouter des séparateurs pour délimiter des familles à l'intérieur de vos entretiens. Les séparateurs sont représentés par des rectangles gris de hauteur fixe placés entre les entretiens, qui mentionnent le "nom de famille" :



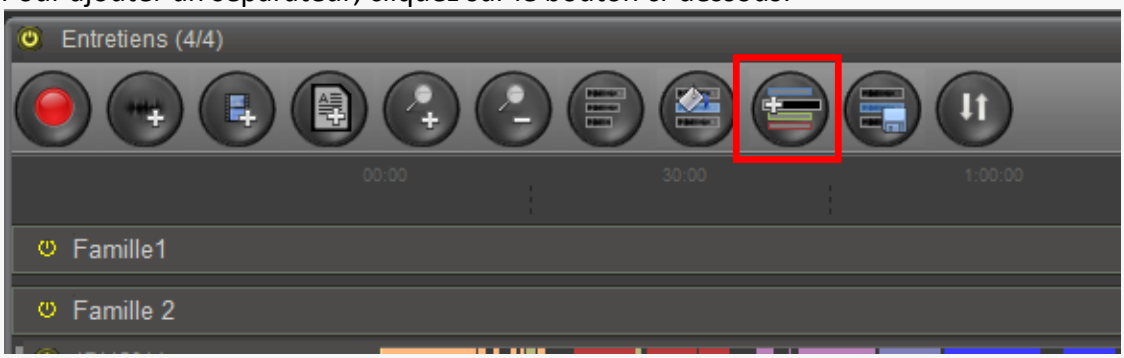
Un entretien placé sous un séparateur appartient à la famille de ce dernier. Lorsque vous ajoutez un document, il est placé d'office dans la dernière famille (c'est-à-dire tout en bas de la liste). Vous pourrez ensuite le déplacer par glisser/déposer.

Ces familles ont essentiellement une fonction d'organisation. On peut en effet "contracter" les familles ou agir collectivement sur les entretiens qu'elles contiennent pour simplifier la représentation du corpus et/ou accélérer les sélections.

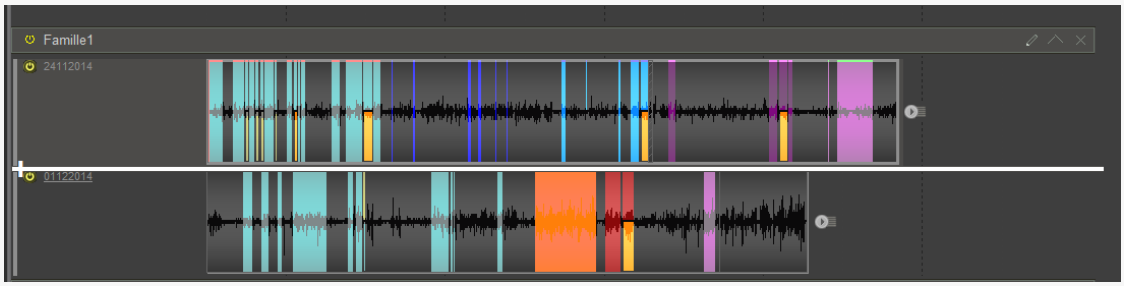


On peut ainsi activer ou désactiver tous les entretiens d'une famille en cliquant sur l'interrupteur collectif placé sur le séparateur.

Pour ajouter un séparateur, cliquez sur le bouton ci-dessous:



Il faut ensuite désigner le premier entretien de la famille. Au survol *du nom* des entretiens, un marqueur blanc simule le placement du futur séparateur:



Cliquez pour ajouter votre séparateur. Renseignez son intitulé et vous aurez une nouvelle famille d'entretiens. Elle se terminera soit au prochain séparateur, soit au dernier entretien du corpus, s'il n'y a pas de séparateur à la suite.

> *Pour vous débarrasser de ce trait blanc sans ajouter de séparateur, recliquez sur le bouton ci-dessus. L'ajout est désactivé*

- Gérer les séparateurs

Vous pouvez renommer vos séparateurs et les supprimer en cliquant sur les boutons ci-dessous :

5.7 Trier vos documents

Lorsque votre corpus s'enrichit, il devient nécessaire de mettre un peu d'ordre pour vous y retrouver. Sonal vous propose plusieurs solutions.

- Les fonctions de tri

Vous avez la possibilité de trier vos documents de différentes manières. Le bouton de tri est placé tout à droite des fonctions au dessus des pistes dans la fenêtre de corpus:

Vous pouvez trier vos documents par ordre alphabétique, par durée, par pourcentage de codage, d'écoute et de durée.

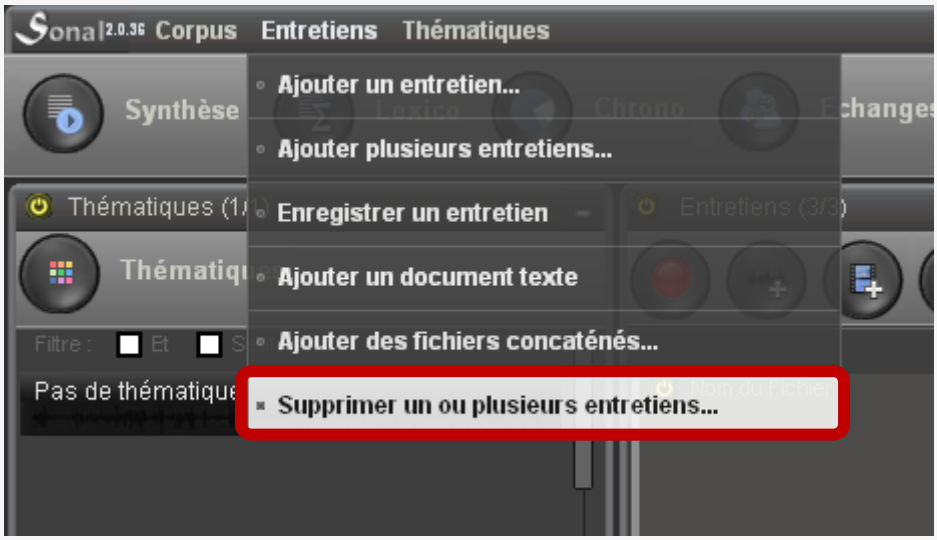
Vous pouvez aussi trier par variables (voir plus bas). C'est-à-dire trier les entretiens sur la base des modalités renseignées dans la base de données (sexe, âge etc…), dans l'ordre où ont été saisies les modalités.
A quoi ça sert? Par exemple, si vous avez plusieurs vagues d'enquête et que vous avez pris soin de mentionner pour chaque document à quelle vague il appartient, vous pouvez faire un tri par variable, choisir la variable "Vague d'enquête" (par exemple). Les entretiens seront triés en fonction de la valeur de modalité pour cette variable et des séparateurs mentionnant chaque modalité seront placés entre les documents. Vous aurez ainsi une famille "Vague1", "Vague2" etc. Si vous voulez ensuite focaliser votre analyse sur une vague en particulier, il vous sera particulièrement facile d'activer ou de désactiver tous les entretiens des différentes vagues pour les analyser.
Cliquez sur le bouton « Variable… » ci-dessous, puis choisissez une variable.

Si vous maintenez cochée la case « ajouter des séparateurs », des séparateurs seront ajoutés automatiquement pour regrouper les documents correspondant aux différentes modalités. *Cela impliquera de supprimer les séparateurs existants*.

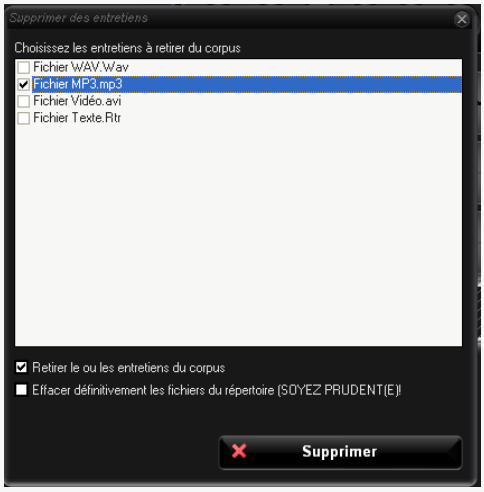
NB: Lorsque des séparateurs sont placés dans la liste des entretiens, le tri se fait à l'intérieur des familles, c'est à dire entre deux séparateurs.

5.8 Supprimer des documents

Pour retirer des documents (son, vidéo ou texte) de votre corpus, utilisez le menu ci-dessous.



La fenêtre qui apparaît vous permet de sélectionner les documents à supprimer.



Par défaut, il vous est seulement proposé de retirer les documents du corpus, c'est-à-dire de laisser intacts tous les fichiers satellites associés au document donné. De la sorte, ces documents pourront être ajoutés à nouveau au projet sans qu'aucune donnée ne soit perdue.

Il est parfois utile de supprimer définitivement tout ou partie des fichiers satellites d'un entretien, par exemple son image (*.BMP), pour la recharger, auquel cas il suffit de cliquer sur « Effacer définitivement les fichiers du répertoire » et de choisir les extensions à supprimer.

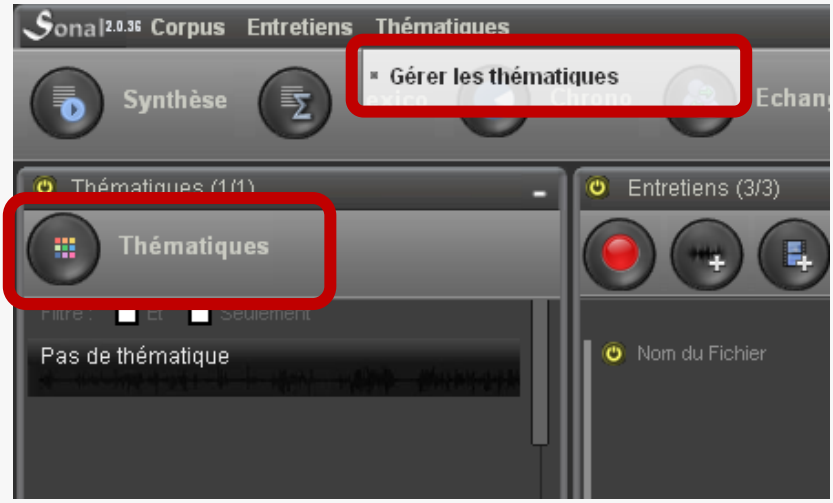
6 Construire la grille thématique

Au sens premier, les thématiques servent à décrire le contenu des échanges. Elles permettront d'apposer des marqueurs de différentes couleurs sur les documents. Dans le cadre d'un corpus d'entretiens semi-directifs, la grille thématique gagnera donc à cadrer avec les différents sujets du guide d'entretien mais on peut imaginer un grand nombre d'usages.

Il ne faut pas perdre de vue que les thématiques n'ont pas qu'une fonction graphique, elles interviennent dans les processus de filtrage et d'analyse. Il faut donc les envisager comme de futurs moyens de distinguer les différents passages. Dans la mesure où il est possible de superposer plusieurs thématiques pour un même extrait, on gagne à faire en sorte que les différentes thématiques s'excluent les unes les autres.

> *Il a 52* ***couleurs disponibles*** *(pour des raisons de transparence) mais les utilisateurs peuvent créer* ***jusqu'à 99 thématiques*** *différentes par corpus.*

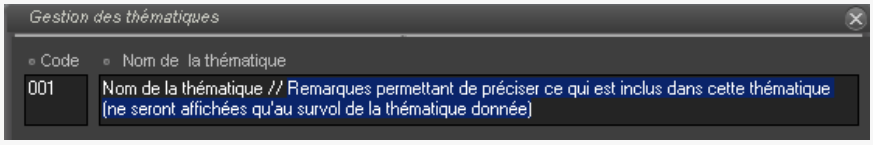
Pour créer des thématiques dans un corpus, il faut soit aller dans le menu « thématiques » puis choisir « gérer les thématiques », soit cliquer sur le bouton « Gérer les thématiques » au sommet du panneau latéral situé à gauche de la page :



6.1 Ajouter une thématique

Dans la fenêtre qui s'ouvre, il faut ensuite **cliquer sur "ajouter une thématique",**

Vous devez d'abord en définir le nom (zone de texte) :



Comme vous pouvez le voir dans l'exemple ci-dessus, vous avez la possibilité de placer des remarques après le séparateur « // ». Ces remarques ne seront visibles qu'au survol.

Puis il faut impérativement choisir une couleur (cliquez sur « choisir une couleur ») :

Puis validez. Votre thématique apparaît maintenant dans la liste. Renouvelez l'opération autant de fois que nécessaire pour créer votre grille de départ.

6.2 Modifier une thématique

Pour modifier une thématique (changer la couleur ou le nom), cliquez sur son nom dans la liste. L'interface d'édition réapparait. Changez ce qui doit l'être puis validez. La modification de la thématique s'applique à tout le corpus

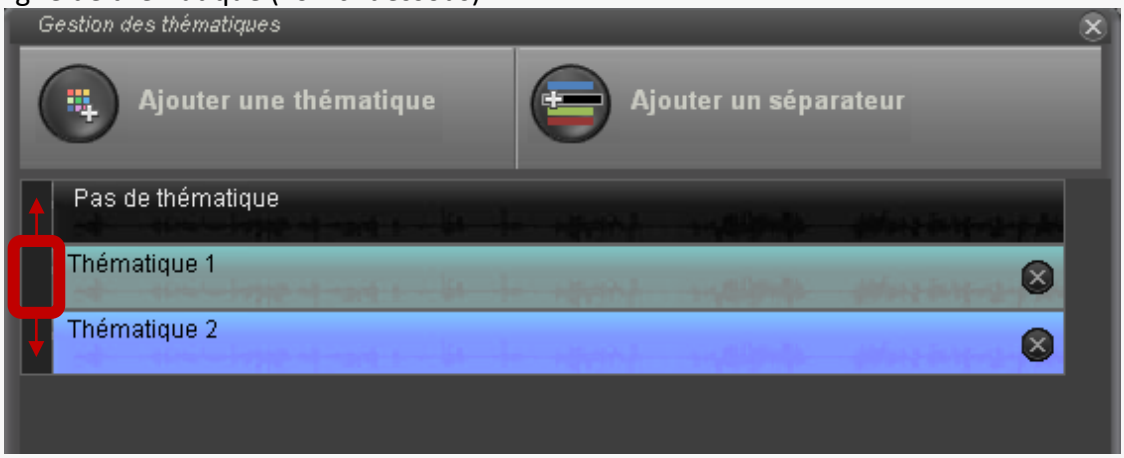
6.3 Supprimer une thématique

Pour supprimer une thématique, cliquez sur la croix située en bout de ligne en face de la thématique considérée. Une confirmation vous est demandée.

> *Avant de supprimer une thématique, il est préférable de s'assurer qu'aucun extrait n'est encodé avec, sinon son code thématique deviendra ininterprétable. Au besoin, commencez par recoder tous les extraits qui la contiennent avant de la supprimer (voir ci-dessous).*

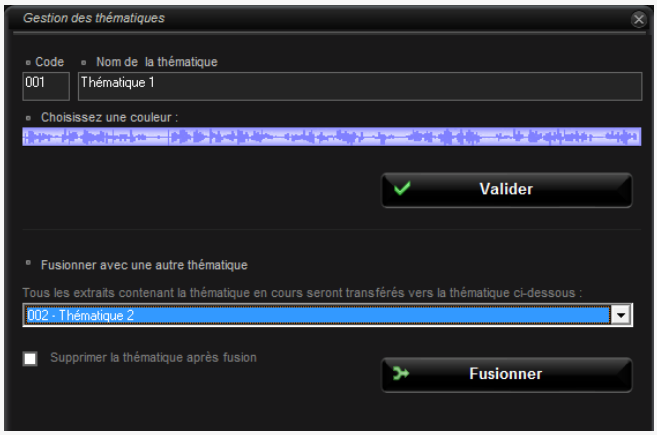
6.4 Déplacer les thématiques

La liste des thématiques est triée par ordre d'ajout ; les thématiques les plus récentes sont donc en bas de liste. Vous pouvez souhaiter les « remonter » ou les descendre afin par exemple de créer des « familles de thématiques » en regroupant toutes celles qui possèdent un couleur proche ou identique. Pour déplacer les thématiques, il faut faire un « glisser-déplacer » en attrapant le rectangle gris situé à gauche de la ligne de thématique (voir ci-dessous).



6.5 Recoder une thématique

Recoder une thématique revient à affecter à tous les extraits qui possèdent la thématique donnée une autre thématique choisie dans la liste des thématiques existantes. Pour cela, on édite la thématique comme si on souhaitait la modifier et l'on choisit, dans la partie basse de la page, la thématique avec laquelle on souhaite « fusionner » la thématique. Dans l'exemple ci-contre, les extraits ayant la thématique 1 se verront affecter la thématique 2.



6.6 Ajouter des séparateurs thématiques

Comme pour les entretiens, il est possible de créer des familles de thématiques afin d'accélérer les éventuelles sélections que l'on peut avoir à opérer.

Pour ajouter un séparateur, il faut cliquer sur le bouton ci-dessous, puis survoler les thématiques. Celle sur laquelle vous cliquerez sera la première de la famille.

6.7 Créer des thématiques en cours de retranscription

La grille thématique peut être enrichie à tout moment depuis les fenêtres d'édition des extraits. Rien n'oblige donc à finaliser la grille avant de commencer à encoder.

Choisissez une thématique
000 - Pas de thématique
001 - Thématique 1
002 - Thématique 2
003 - Thématique 3
004 - Thématique 4
Choisissez une thématique
Code
Nom de la thématique
005
Nom de la thématique
Choisissez une couleur :
Ajouter une thématique

7 Construire et renseigner la base de données

La base de données permet de stocker des informations au sujet des documents du corpus.

> *Bien qu'il ne soit pas absolument indispensable de s'en servir lorsque l'on travaille dans Sonal, son intérêt se fait de plus en plus évident à mesure que le corpus s'enrichit : elle offre une vision d'ensemble des personnes interrogées par exemple, et permet des filtrages qui facilitent la recherche et l'analyse des matériaux. Plus fondamentalement, la définition des variables clés au sujet des répondants pousse le chercheur à formaliser les catégories analytiques importantes de son enquête, ce qui ne peut être que bénéfique.*

L'utilisateur définit jusqu'à 99 variables, auxquelles il pourra affecter 98 modalités (99=non réponse). Les variables typiques d'une enquête sociologique sont les incontournables : Sexe, Âge, Pcs, etc. mais l'utilisateur est totalement libre d'encoder ce qu'il veut.
Pour chaque entretien, on pourra ensuite renseigner les informations demandées (Homme, 50 ans, Employé…), soit en les tapant directement au clavier, soit en choisissant la modalité dans une liste constituée des modalités déjà choisies.

La base de données vise également à éprouver des hypothèses de recherche : si vous pensez qu'un attribut particulier est susceptible d'influencer les discours et représentations de vos interviewés, intégrer cette variable dans la base de données vous permettra par la suite de vérifier si cet attribut a bel et bien une influence.

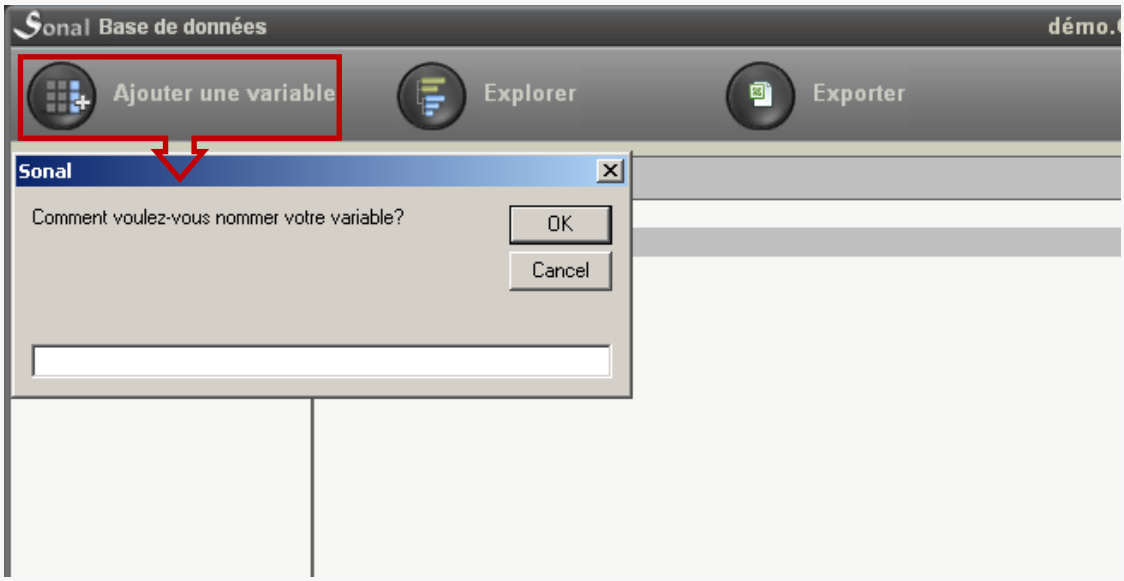
> *Même si ce travail peut parfois sembler très formel, il ne faut pas oublier que la connaissance précise des personnes interrogées décroît avec le temps et le nombre, et que le fait de garder de manière formalisée des informations clés au sujet des répondants présente l'intérêt par la suite d'éviter de se perdre dans ses données. Pour peu que le corpus commence à contenir plus d'une quinzaine d'entretiens, il devient utile de pouvoir éditer rapidement des tableaux récapitulatifs sur la population enquêtée sans avoir à compter à la main les entretiens répondant à telle ou telle caractéristique.*

On accède à la base de données en cliquant sur le bouton « Base de données » dans la page de corpus :

En l'absence de variables, la base de données a cette apparence :

7.1 Ajouter une variable

La première chose à faire est d'ajouter une variable, en cliquant sur « ajouter une variable ». Une boite de dialogue apparaît, qui vous demande le nom (en toutes lettres de cette variable). Renseignez-le (exemple : Sexe ou Age), puis validez.

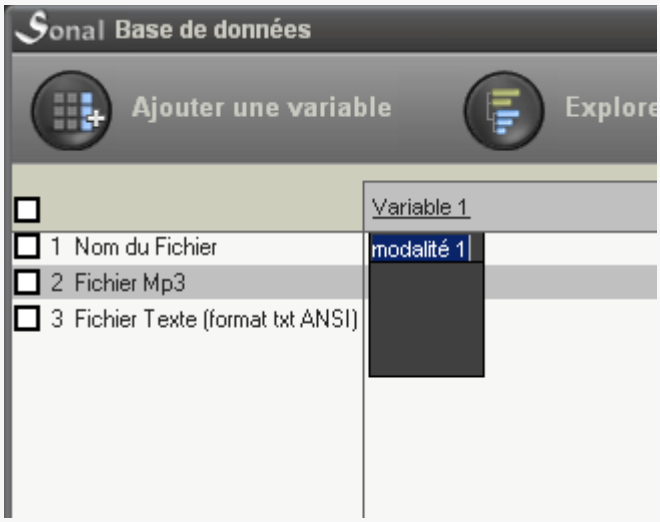


La variable ajoutée apparaît dans la base :

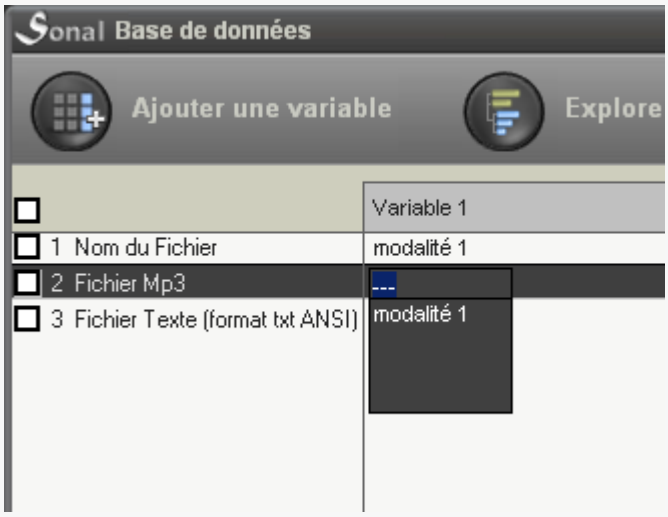
Dans la mesure où aucune modalité n'est renseignée pour les différents documents du corpus, on trouve « --- » à l'intersection des lignes et des colonnes.

7.2 Renseigner une variable

Le renseignement des variables – i.e. le choix d'une modalité pour chaque entretien – se fait de manière très simple. Cliquez sur la case à modifier, un menu d'abord vierge apparaît. Saisissez la valeur de modalité puis **validez avec « entrée ».**



Cette modalité sera par la suite disponible à la sélection pour les autres documents.

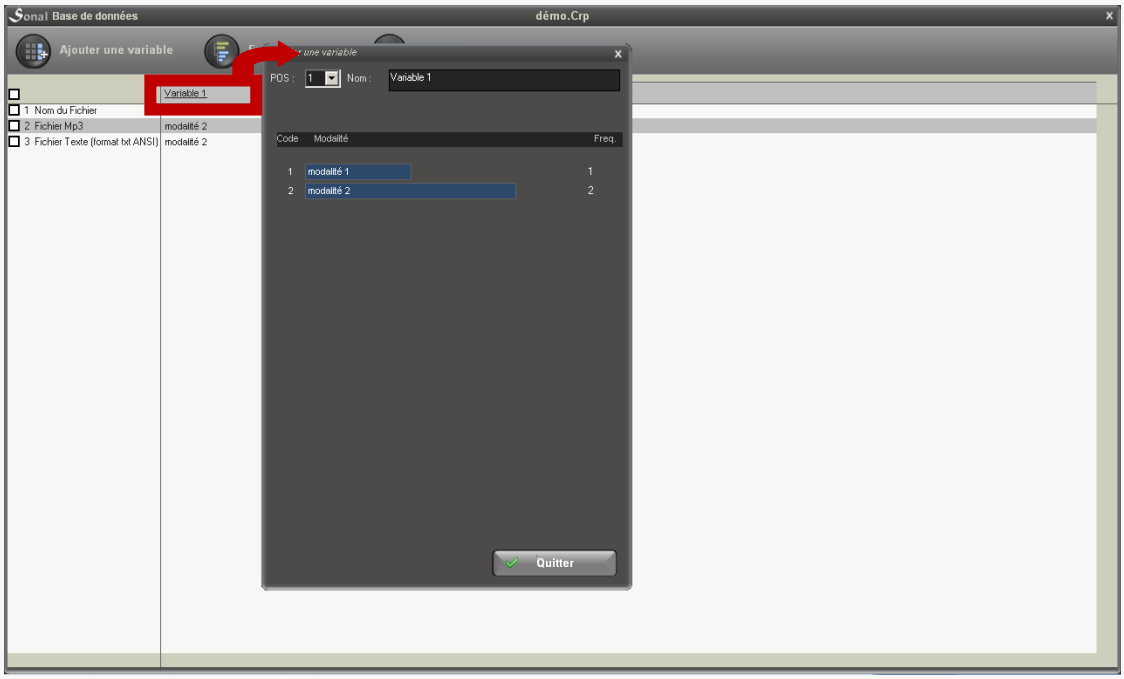


Vous construisez ainsi le dictionnaire de la base de données.

Les modalités apparaitront ensuite toujours ***dans l'ordre où elles ont été rentrées.*** Si l'ordre a une importance (par exemple pour des classes d'âge etc.), la méthode la plus simple consistera, pour le premier document de votre corpus, à renseigner successivement toutes les modalités dans le bon ordre. Par exemple à taper toutes les classes d'âge dans l'ordre, en validant à chaque fois, puis à choisir la bonne dans la liste.

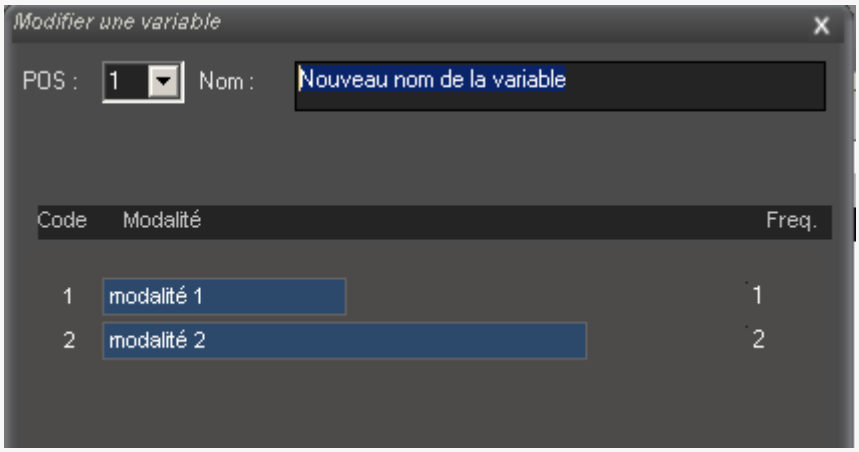
7.3 Modifier une variable ou une modalité

Pour modifier une variable, cliquez sur son intitulé :

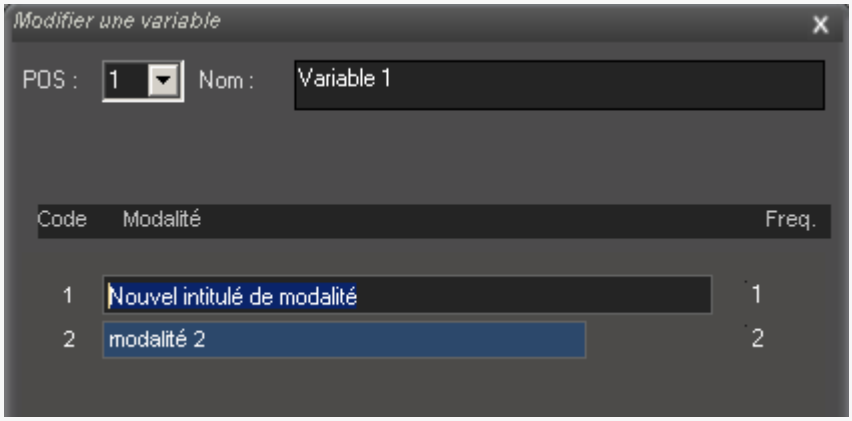


Dans la fenêtre de tri à plat qui apparaît alors, vous pouvez modifier :

- l'intitulé de la variable :



- L'intitulé les modalités (cliquez dessus, tapez le nouveau nom, et *validez avec entrée*):



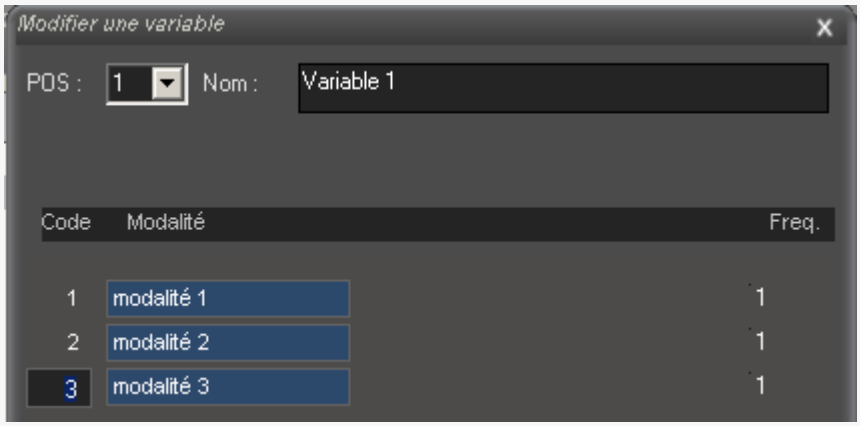
Pour **supprimer** une modalité non utilisée, effacez son contenu, puis validez.

7.4 Recoder des modalités

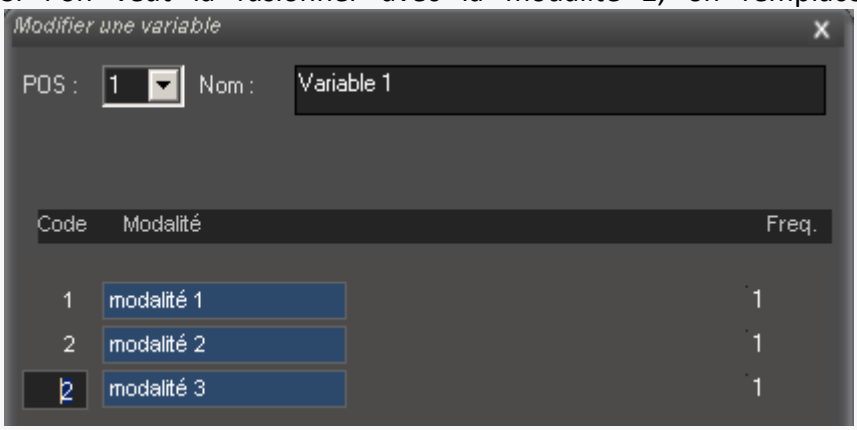
Pour **recoder** une modalité (par exemple regrouper plusieurs réponses), modifiez son code puis validez. Si le nouveau code choisi pour une modalité existe déjà, Sonal vous proposera de fusionner les deux modalités. Vous pourrez ainsi regrouper des réponses en affectant successivement plusieurs modalités à un même code.

Voici par exemple comment fusionner les modalités 2 et 3 de notre variable 1 :

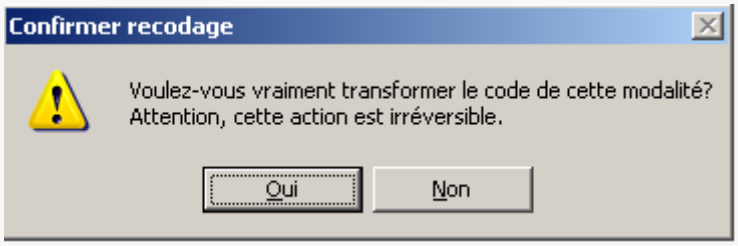
1) On clique sur le code de la modalité 3 :



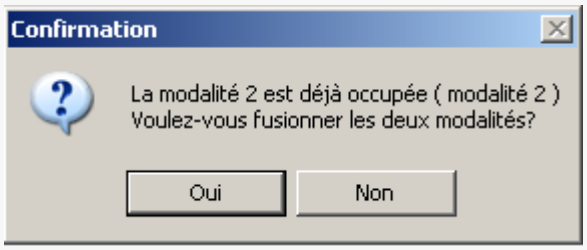
Si l'on veut la fusionner avec la modalité 2, on remplace le 3 par un 2 :



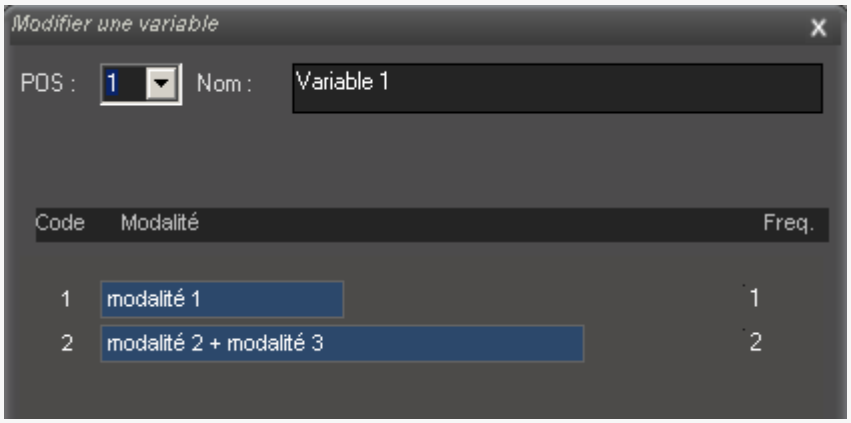
Puis on valide avec entrée. Un premier message vous alerte que vous allez transformer le dictionnaire:



Un second message vous demande confirmation avant de fusionner deux modalités :



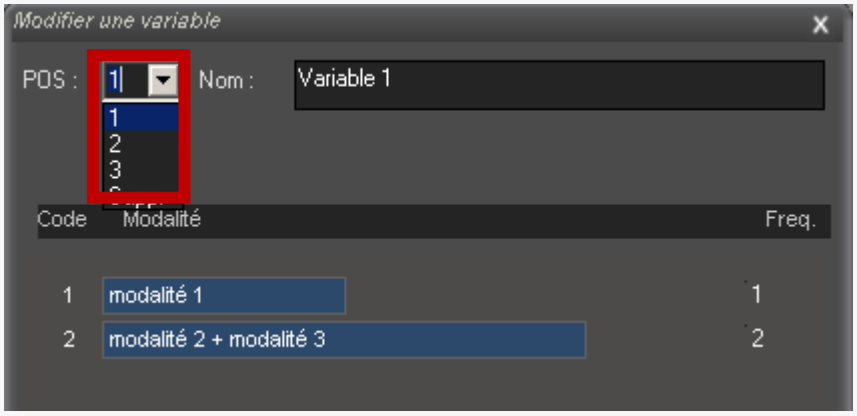
Répondez oui, vous obtenez alors le tri à plat suivant :



L'intitulé de la modalité 2 rappelle la fusion que vous venez d'opérer et le tri à plat tient compte du regroupement.

7.5 Déplacer les variables

Pour déplacer les variables vers la droite et la gauche, utilisez le menu déroulant qui vous indique la position de la variable :

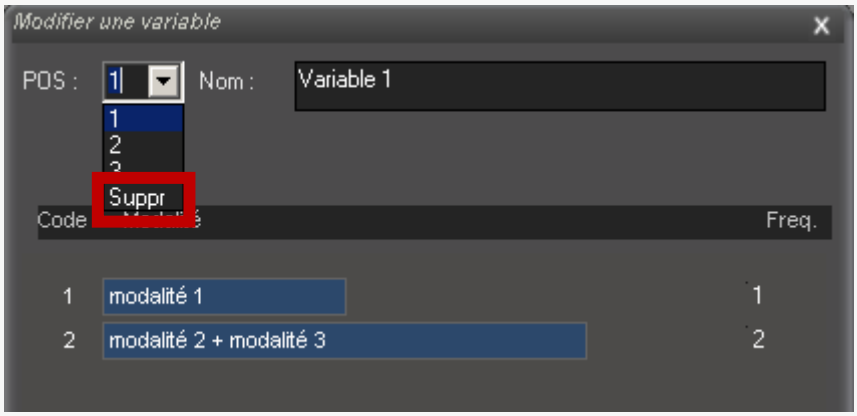


Vous pouvez aussi utiliser la méthode suivante : faites un **clic droit** sur l'entête de colonne de la variable puis **déplacez là** jusqu'à la positionner sur la colonne où vous souhaitez la placer.

NB : Déplacer une variable va consister d'une part à modifier le fichier corpus et tous les fichiers RTR satellites *référencés dans le corpus*. La modification ne s'appliquera pas d'elle-même à tous les entretiens présents dans le dossier du projet.

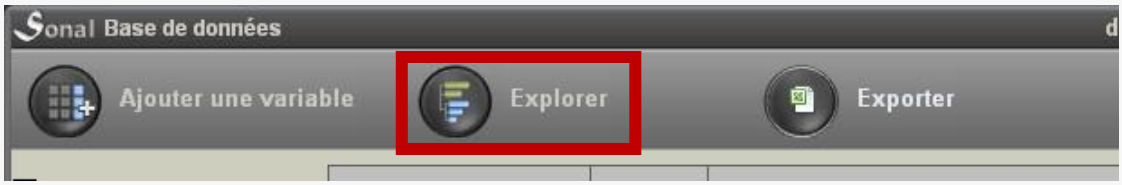
7.6 Supprimer des variables

Pour totalement supprimer une variable, choisissez « Suppr » dans le menu déroulant.



7.7 Explorer la base de données

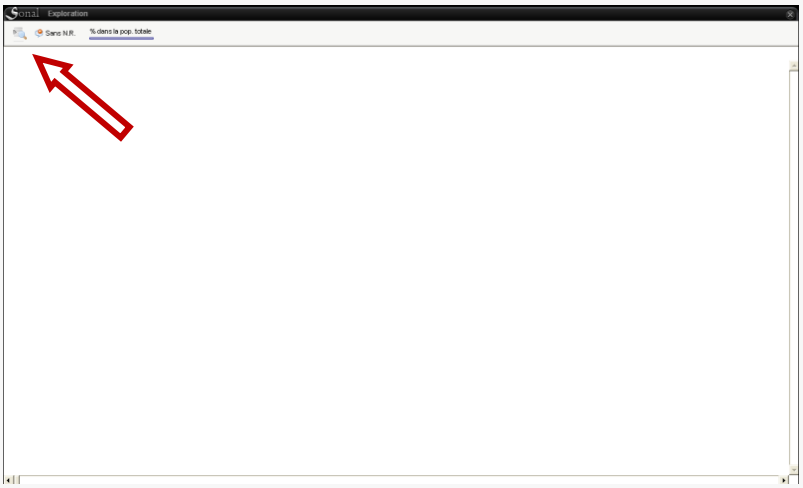
Pour analyser la composition de votre corpus, vous pouvez utiliser le bouton « explorer » situé dans la fenêtre où figure la base de données :



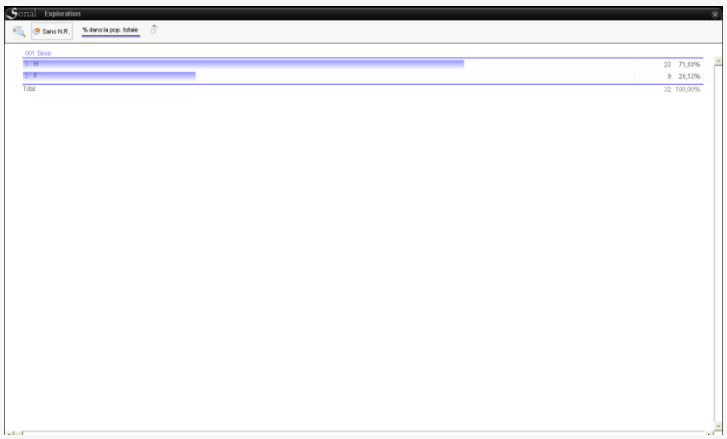
Lorsque vous faites cela, Sonal crée une base de données au format Trideux dans un dossier nommé « Trideux » qui est situé dans le dossier de votre projet. Cette base porte le nom de votre corpus suivi de « (DAT) ».

Depuis la version 1.9, Sonal va directement lire dans cette base sans qu'il vous soit nécessaire de lancer Trideux.

L'interface a alors cette apparence :

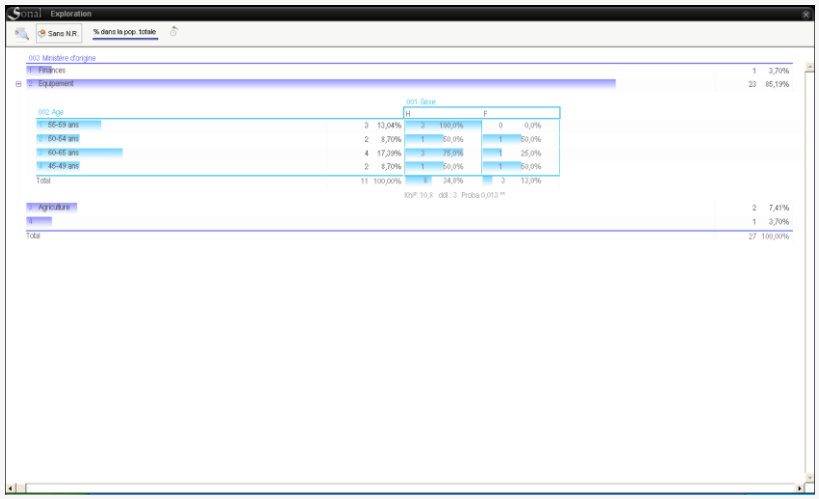


En cliquant sur le bouton qui est situé en haut à gauche de la page, vous aurez accès à la liste des variables de la base de données. En choisissant une variable, vous aurez accès à son tri à plat :



L'intérêt de la fenêtre d'exploration est qu'il vous sera possible de **cliquer sur l'une des modalités pour avoir un nouveau tri à plat**, mais portant uniquement sur la population en question. Vous pouvez ainsi « creuser » dans votre base à volonté :

La fenêtre d'exploration vous permet en quelque sorte de faire des « tris croisés à plat ». Mais vous pouvez également y réaliser des tris croisés classiques, en cliquant - au survol d'une variable -, sur le bouton situé en bout de ligne :



Si vous voulez utiliser des outils plus poussés pour vos analyses, vous pouvez évidemment ouvrir votre base directement dans Trideux.

7.8 Exporter la base de données

L'un des intérêts qu'il y a à renseigner une base de données dans Sonal est que le logiciel vous offrira la possibilité, en fin ou en cours d'enquête, de faire très simplement le point sur la composition sur votre corpus : combien avez-vous interrogé de personnes de telle ou telle catégorie ? etc. Ce travail d'inventaire est facile à réaliser sur une petite quantité d'entretien ; il devient nettement plus fastidieux à mesure que votre corpus grossit.

Sonal vous permet de faire plusieurs choses.

Vous pouvez d'abord exporter votre base vers un tableur. Pour cela, cliquez sur « Exporter ».

Sonal vous créera alors un fichier tabulé lisible par Excel ou Calc qui reprendra, pour chaque ligne, les informations contenues dans la base de données.

Encoder des documents

- Qu'appelle-t-on encodage dans Sonal ?

Vous avez importé vos premiers enregistrements sonores ou vidéo, ou vos premiers textes. Vous souhaitez maintenant commencer à les encoder.

- Dans Sonal, le travail d'encodage des **enregistrements sonores ou vidéo** implique plusieurs tâches :

  - ✓ Segmenter les documents (créer des "extraits" audio/vidéo ou textuels)
  - ✓ Affecter des thématiques et des mots-clés à ces extraits
  - ✓ gérer le contenu textuel des extraits (retranscrire, annoter)

Le travail d'encodage peut être organisé de plusieurs manières, selon vos besoins et vos matériaux. Vous disposez d'une grande souplesse dans l'organisation des opérations.

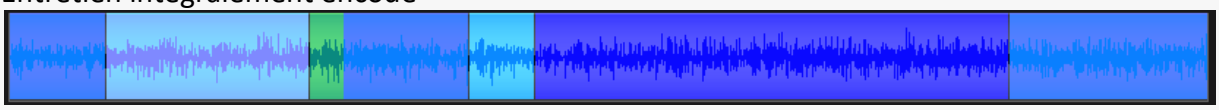
Sonal n'impose aucune méthode de travail aux utilisateurs. Vous pouvez commencer par saisir le texte (retranscription), sans vous soucier du découpage, - car vous pourrez le faire ensuite -, soit choisir de commencer par découper la bande, en mettant éventuellement quelques remarques, avant de retranscrire le contenu des extraits si vous en avez l'usage.

Vous pouvez donc avoir trois orientations différentes.

- Soit vous ne vous occupez d'abord que des retranscriptions.
- soit vous ne vous occupez d'abord que du découpage des extraits sonores/vidéos,
- soit vous vous occupez des deux en même temps.

Vous êtes également libre de choisir d'opter pour un encodage continu ou discontinu. Soit vous surlignez les passages intéressants, soit vous procédez à un encodage *continu* qui reviendra à chapitrer la bande sans laisser de passage vierge.

Entretien intégralement encodé



Entretien partiellement encodé

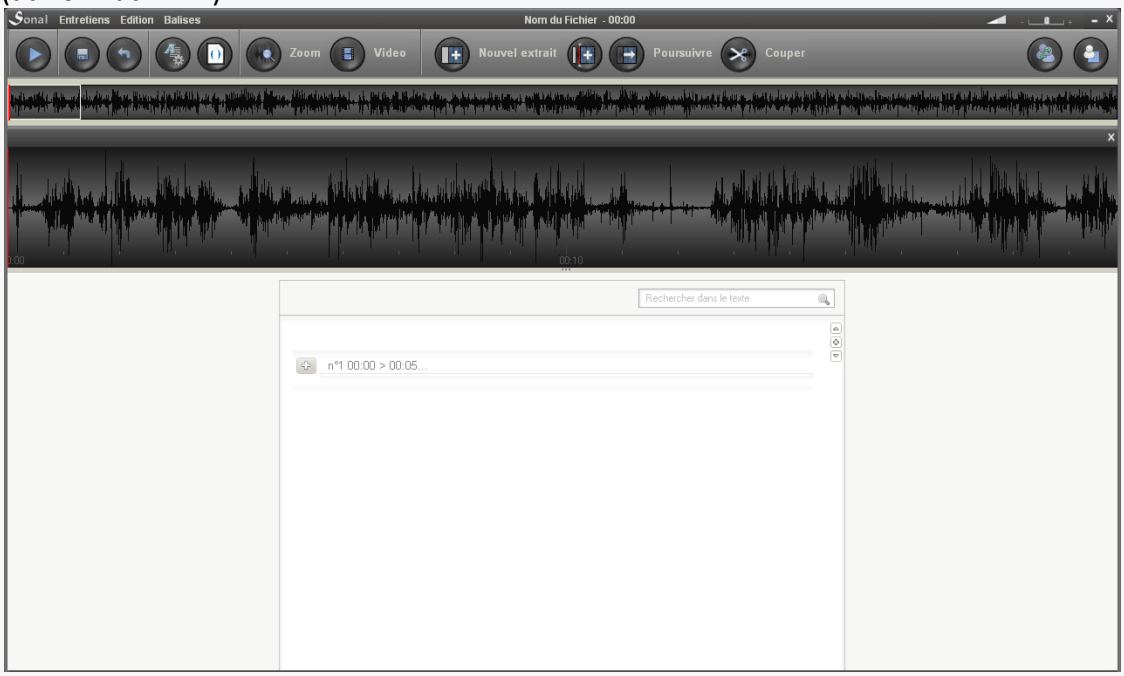
Attention, si vous utilisez Sonal dans une perspective analytique, l'encodage intégral s'impose. En effet, il faut encoder intégralement le son pour procéder à des analyses chronométriques, et retranscrire intégralement le texte pour procéder à des analyses lexicométriques.

8 Coder des enregistrements audio/vidéo

Dans Sonal 2, tout le travail de découpage et de retranscription des *enregistrements* audio et/ou vidéo se faire dans une fenêtre spécifique, appelée "fenêtre de codage" (pour les documents textuels, voir ici). Cette fenêtre est spécifiquement pensée pour rendre constamment possible la prise en charge conjointe du texte et du son ou de la vidéo.

Elle affiche à la fois le son d'origine, les extraits codés et le texte retranscrit. On y accède via le petit bouton en bout de ligne en face de la piste (1) ou, après avoir cliqué *sur le nom du document*, sur le bouton (2) qui apparaît dans le panneau latéral :

La fenêtre de codage a cette apparence lors de la première ouverture d'un entretien (au format WAV):

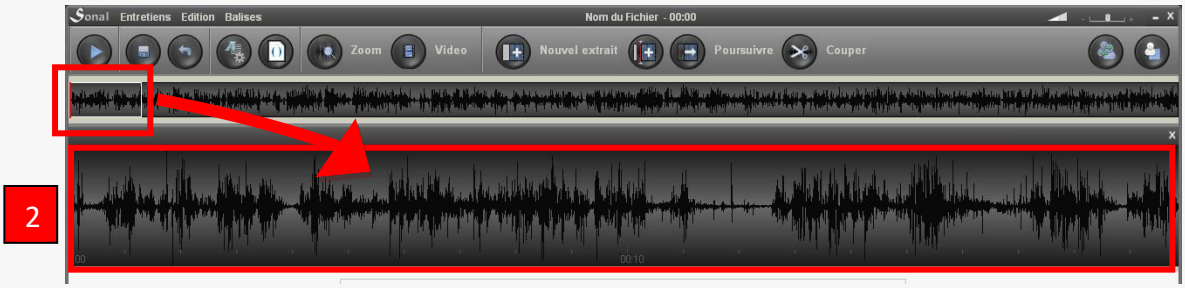


- Généralités

La fenêtre est divisée en trois parties :

- Dans la partie supérieure, l'intégralité de la bande (1), représentée par le spectre sonore (si le fichier est au format WAV) :

- Juste en dessous, une vue zoomée (2) de la partie qui est encadrée par un rectangle blanc dans la partie supérieure. Cette vue zoomée permet une meilleure vision des extraits et des points de synchronisation.



Cette « barre de zoom » s'affiche ou se dissimule en utilisant le bouton « zoom » :

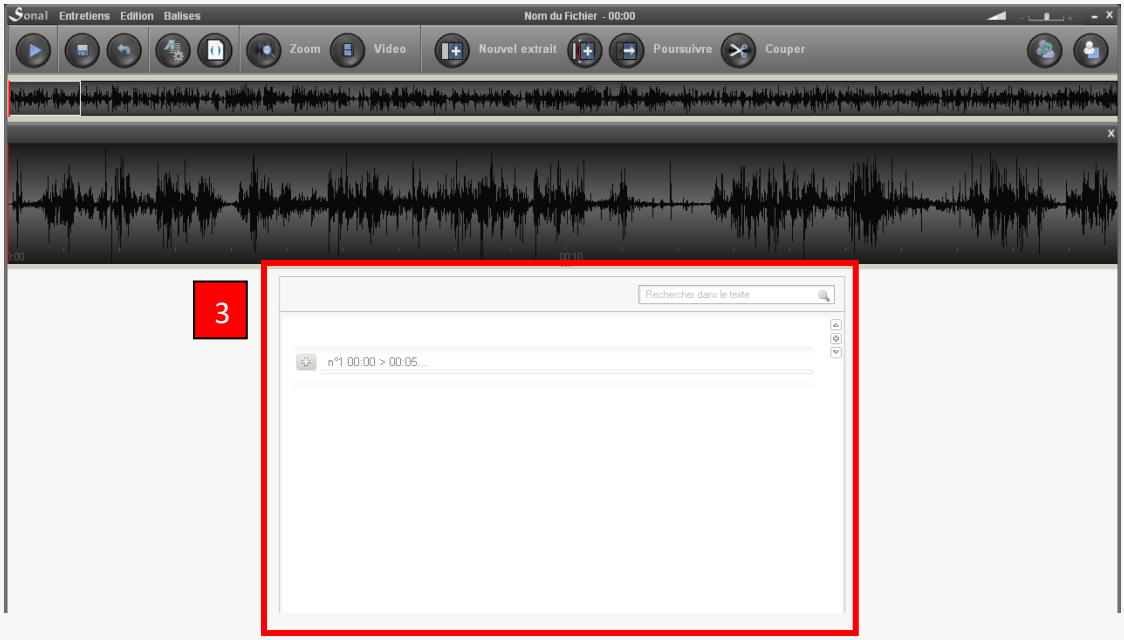
Par défaut, le zoom représente deux minutes de bande, mais on peut augmenter ou diminuer ce temps en *dessinant* à la souris la portion que l'on veut voir zoomée.

> *Dans la vue totale de la bande, enfoncez la souris au début de la portion que vous voulez zoomer, maintenez enfoncé et décalez-vous vers la droite. L'endroit où vous relâcherez la souris délimitera la fin de la portion*

On peut également utiliser les boutons + et – qui apparaissent tout à droite de la barre de zoom lorsqu'on la survole :

Conseil : il n'est pas recommandé de délimiter des portions trop longues. Outre que cela ne sert pas à grand-chose, le temps d'affichage risque d'être important.

- Enfin, dans la partie inférieure de la page, sur fond blanc, la représentation verticale des extraits et de leur contenu textuel (3):



Au premier chargement d'un fichier, cette zone est logiquement vierge, à l'exception d'un premier extrait « fantôme » (voir plus bas).

Lorsqu'un entretien a été codé, la fenêtre a cette apparence :

Le texte et le son sont synchronisés. En cliquant sur un mot on réentend le passage considéré et réciproquement.

- Contrôler de la lecture de la bande audio/vidéo dans la fenêtre de codage

Pour lancer ou arrêter la lecture, vous pouvez utiliser le bouton « play » en haut à gauche de la fenêtre, mais aussi la **barre d'espace.**

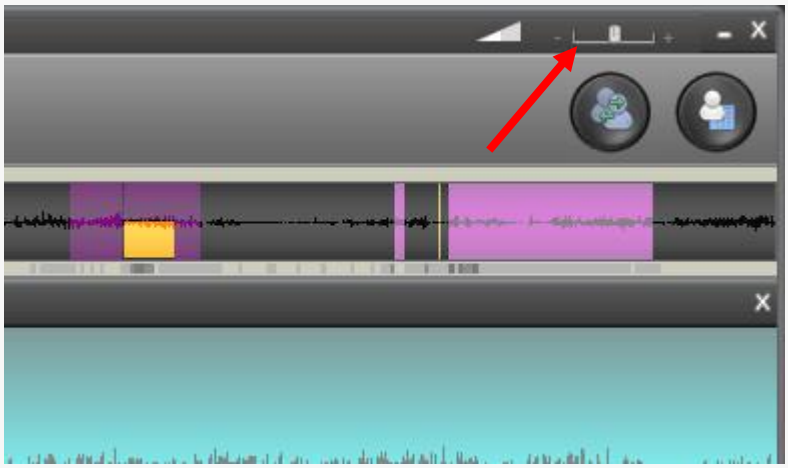
En utilisant la souris, vous pouvez repositionner la lecture.

Si vous cliquez dans la partie supérieure de la fenêtre, la portion zoomée (matérialisée par le rectangle blanc) va se déplacer pour débuter quelques secondes avant là où vous avez cliqué.

Si vous cliquez dans la partie zoomée, la lecture se positionnera là où vous avez cliqué.

- Accélérer ou ralentir la lecture

Pour modifier la vitesse de lecture de la bande, utilisez la petite réglette que vous avez en haut à droite de la page (voir ci-dessous).



Le résultat est un peu différent selon que vous utilisez un fichier WAV ou MP3. Le WAV va "pitcher", transformant la hauteur de voix de vos locuteurs (ce qui donne des résultats parfois bizarres), le MP3 va seulement accélérer ou ralentir. Sur ce plan, net avantage au MP3 donc...

Créer des extraits

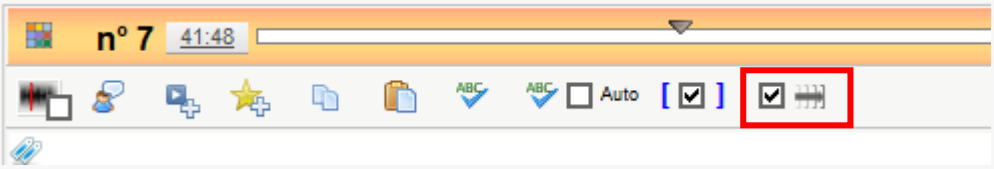
Quoi que vous souhaitiez faire avec Sonal, le **découpage des extraits est une tâche incontournable**. Même si vous souhaitez retranscrire intégralement un entretien sans vous soucier de découpage thématique, il vous faudra *a minima* créer un premier extrait qui contiendra tout le texte de l'entretien.

- Généralités sur les extraits

Dans Sonal, lorsqu'on code un fichier son ou vidéo, **créer un extrait veut dire « définir la position de début d'un nouvel extrait ».** On ne précise pas la position de fin car, d'une manière générale, Sonal est conçu dans l'idée que lorsqu'on écoute un enregistrement pour l'encoder, on ne sait pas encore quand se terminera un extrait lorsqu'on le débute. C'est en écoutant ou en regardant qu'on définira la fin de l'extrait. Bien sûr, on peut facilement renseigner à la main les coordonnées des extraits si l'on sait déjà quand on souhaite qu'ils se terminent, mais par défaut, le fonctionnement est le suivant :

Lorsque l'on crée un extrait, celui-ci dure par défaut 5 secondes. On passe en mode édition et **l'extrait <u>s'allonge au fur et à mesure de l'écoute, jusqu'à ce qu'on arrête la bande ou qu'on valide l'extrait</u>**. Concrètement, la bordure de droite va se déplacer avec le lecteur pour signifier l'allongement de l'extrait.

Une fois l'extrait validé, **si on l'ouvre à nouveau**, **celui-ci ne s'allongera plus** au fil de la lecture, **la bande s'arrêtera à la fin de l'extrait**. Pour faire en sorte que l'extrait s'allonge à nouveau, cliquez sur la case ci-dessous.



Il y a trois manières de créer des extraits. Soit vous créez un extrait à la suite du dernier extrait (1), soit vous créez un extrait à la position du lecteur (2), soit vous créez un extrait à la souris (3).

- 1) Créer un extrait au point le plus avancé

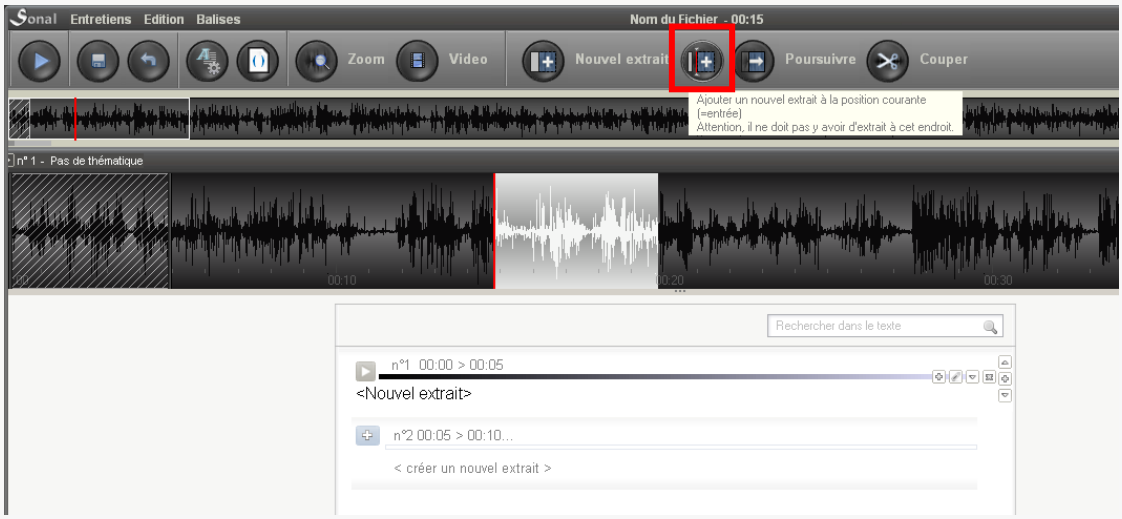
Le(s) bouton(s) « nouvel extrait » (voir ci-dessous), sont particulièrement adaptés à un encodage continu (tous les extraits se suivent et sont accolés). Cliquer sur ce bouton implique tout simplement de **démarrer un nouvel extrait immédiatement à la suite de l'extrait le plus avancé**. S'il n'y a pas encore d'extrait, ce sera donc au tout début de la bande. Lorsque l'on survole ces boutons, Sonal représente l'extrait à créer dans la vue zoomée (1).

Une fois que l'on a cliqué sur ce bouton, sonal crée un premier extrait de 5 secondes à la position donnée et l'ouvre et lance la bande pour que vous puissiez commencer à l'éditer :

L'édition des extraits est présentée plus bas.

- 2) Créer un extrait à la position du lecteur

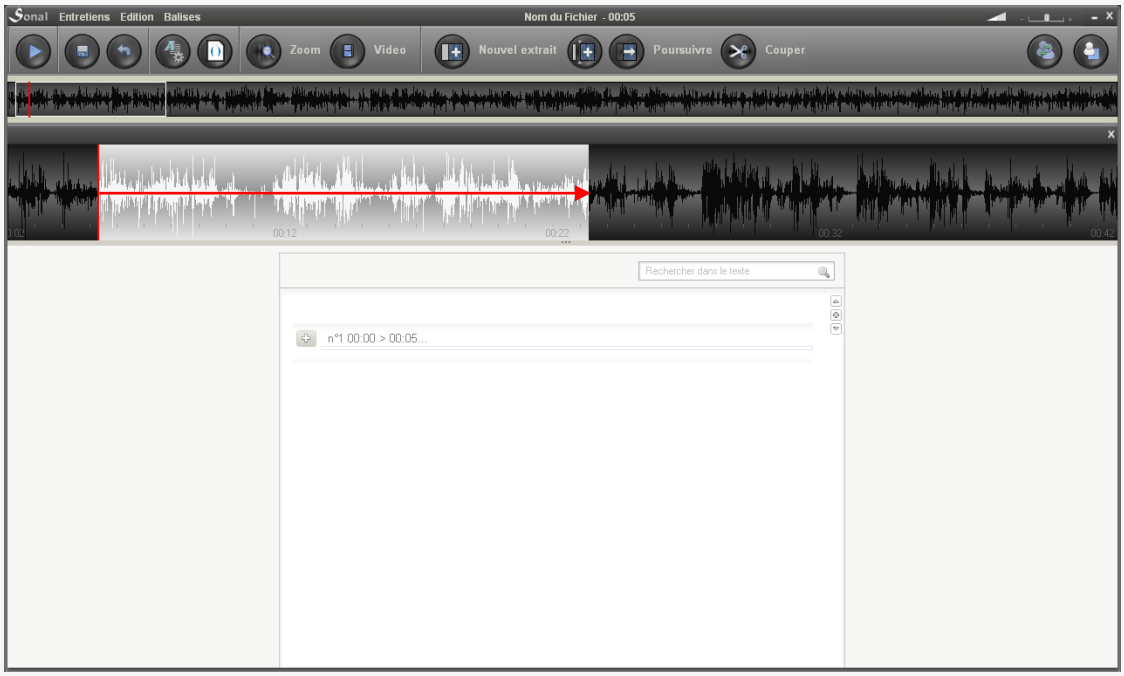
Dans la mesure où l'on ne souhaite pas toujours créer un extrait au tout début de la bande, où immédiatement à la suite du précédent, il est également possible de **créer un extrait à la position atteinte par le lecteur audio**. En pareil cas, il faut positionner le lecteur audio à l'endroit voulu et cliquer sur le bouton ci-dessous. On peut également faire « entrée » au clavier en cours de lecture :



Important : Dans Sonal, le codage est ainsi fait qu'il n'est pas prévu que les extraits se superposent. Bien qu'il existe des moyens indirects d'arriver à un tel résultat, le plus souvent, Sonal tentera de vous empêcher de créer des extraits superposés. Si vous l'avez fait involontairement, un système de débogage vous permettra de corriger le codage.

- 3) Créer un extrait à la souris.

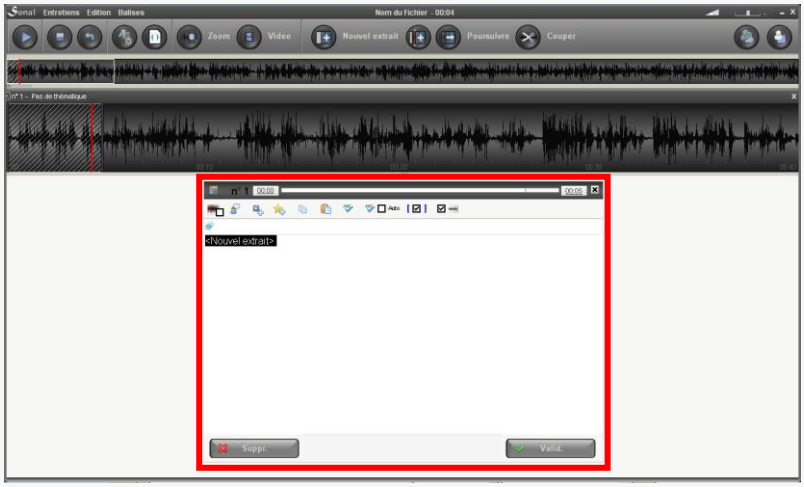
Pour créer manuellement un extrait, il faut dessiner la zone correspondante dans la vue zoomée de la bande. Sélectionnez le début en enfonçant le bouton gauche de la souris, maintenez-le enfoncé et faites glisser votre souris vers la droite ; relâchez le bouton à la fin de la portion que vous voulez sélectionner.



Une fois la partie sélectionnée, relâchez la souris. L'extrait est créé et Sonal l'ouvre pour que vous éditiez son contenu (voir plus bas). Vous pouvez cliquer sur le bouton « valider » en bas à droite de la page.

8.1 Editer, valider les extraits

Lorsque l'on crée un nouvel extrait ou qu'on « ouvre » à nouveau un extrait existant pour l'éditer, une fenêtre d'édition se superpose à la zone de lecture :

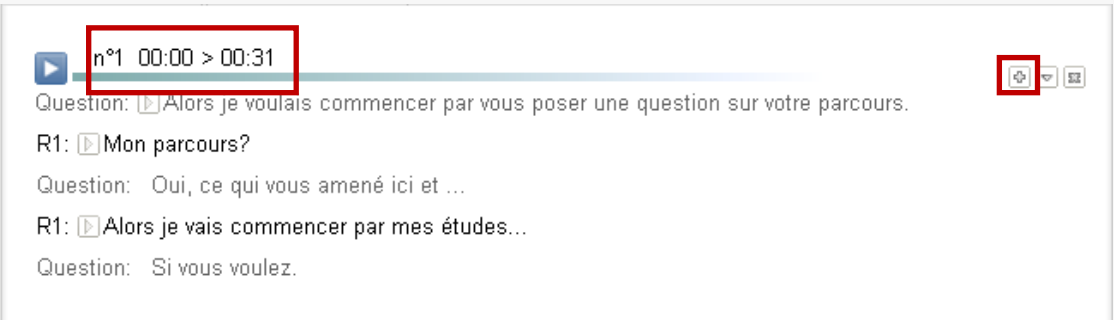


La zone qui apparaît alors vous permet de modifier le contenu des extraits, c'est-à-dire à la fois de définir leurs coordonnées, de leur affecter une thématique (code couleur), des mots clés, et bien sûr de retranscrire leur contenu.

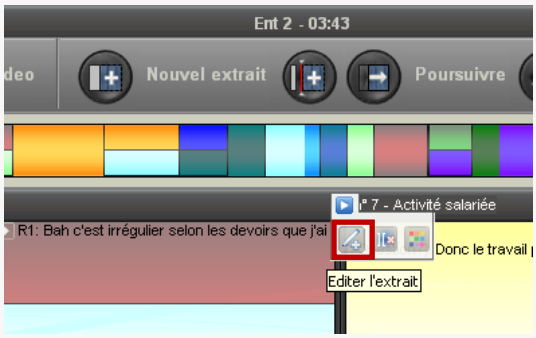
Si les nouveaux extraits sont immédiatement proposés à l'édition, il faut également savoir les rouvrir pour modification.

- Rouvrir et fermer un extrait

Si vous souhaitez rouvrir un extrait existant dans la page de codage, vous pouvez cliquer sur les éléments ci-dessous dans les entêtes d'extraits.



Autre méthode : après *clic-droit* sur les boutons play qui marquent les débuts d'extrait dans la vue zoomée, vous pouvez choisir le bouton ci-dessous dans le menu qui apparaît:



- Rouvrir le dernier extrait

Cas particulier. Si vous souhaitez rouvrir un extrait **pour continuer à le retranscrire ou l'allonger**, vous pouvez utiliser le bouton « poursuivre ».

Sonal rouvrira l'extrait le plus avancé et se placera à la fin de celui-ci.

La validation d'un extrait se fait via le bouton « valid. » en bas à droite de la zone. Pour sortir de l'extrait sans enregistrer les modifications, cliquez sur la croix en haut à droite. L'extrait est refermé sans sauvegarde des modifications.

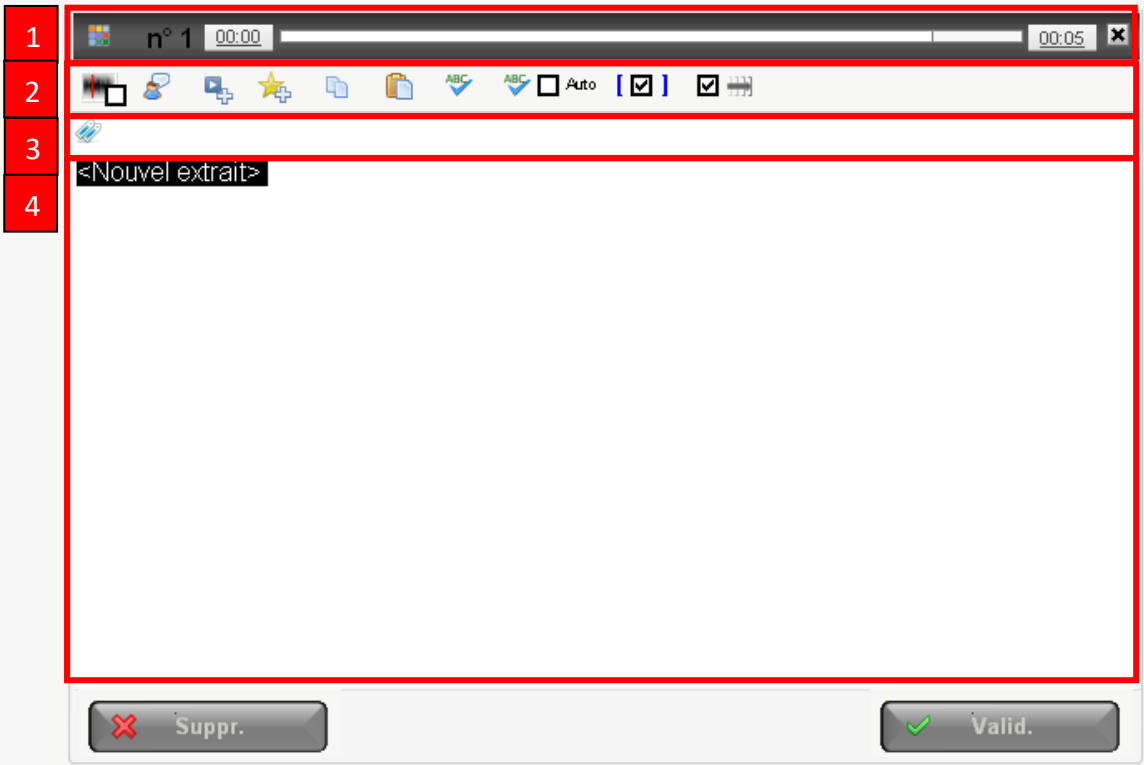
- Quand valider les extraits ?

Il est largement préférable de découper les documents au fur et à mesure en validant au moins quelques extraits. **Un seul extrait très long n'est pas une bonne chose** car cela ralentit le logiciel. Pour donner un ordre d'idées, il ne faut pas que vos extraits durent plus d'un quart d'heure.

- Propriétés de la zone d'édition

Voici plus en détail la manière dont fonctionne la zone d'édition des extraits.

Il y a 4 espaces dans cette fenêtre :
La partie 1 permet de contrôler la thématique, les coordonnées de l'extrait et la position du lecteur.
La partie 2 contient les options de retranscription et notamment le mode dictée.
La partie 3 contient les mots clés (tags).
La zone de texte 4 contient le texte associé à l'extrait.



Contrôle de la bande en mode édition.

Dans la mesure où cette zone est dévolue à la saisie au clavier, la barre d'espace retrouve sa vocation initiale et ne contrôle plus la bande.

Il y a cependant des astuces :

1 – Si vous tapez **trois fois de suite espace**, Sonal rendra la « main » à la barre d'espace pour contrôler la bande : la barre d'espace n'ajoutera plus d'espace dans le texte, elle servira seulement à piloter le lecteur.

2 – En utilisant la combinaison **Alt + espace**, vous pouvez arrêter ou démarrer la lecture.

8.2 Définir manuellement les coordonnées d'un extrait

Si le fonctionnement de base de Sonal suppose un allongement progressif des extraits au fur et à mesure de l'écoute, l'utilisateur dispose de différents moyens pour définir manuellement les coordonnées d'un extrait.

Il faut d'abord savoir que toutes les bordures d'extraits qui figurent dans la barre de zoom peuvent être déplacées manuellement vers la gauche ou la droite (pour peu qu'elles ne se superposent pas avec un extrait existant). Des ajustements marginaux peuvent donc être opérés à la souris.

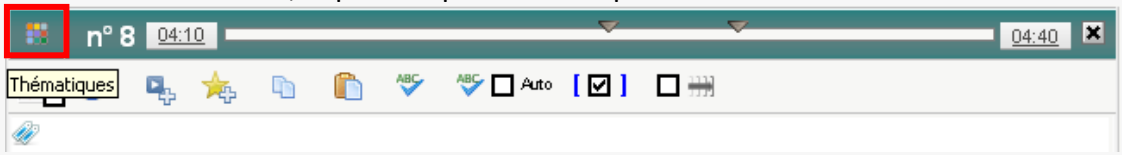
On peut également avoir une approche plus fine en cliquant sur les coordonnées de début ou de fin d'extrait pour faire apparaître une fenêtre dédiée :

Vous pouvez alors renseigner au clavier les coordonnées (soit en secondes soit en h:m:s).
Le bouton 1 cale le début de l'extrait sur la fin de précédent ou, à défaut, au début de l'enregistrement. Le bouton 2 cale la fin de l'extrait au début du suivant ou, à défaut, à la fin de la bande.

8.3 Affecter une ou des thématique(s) à un extrait

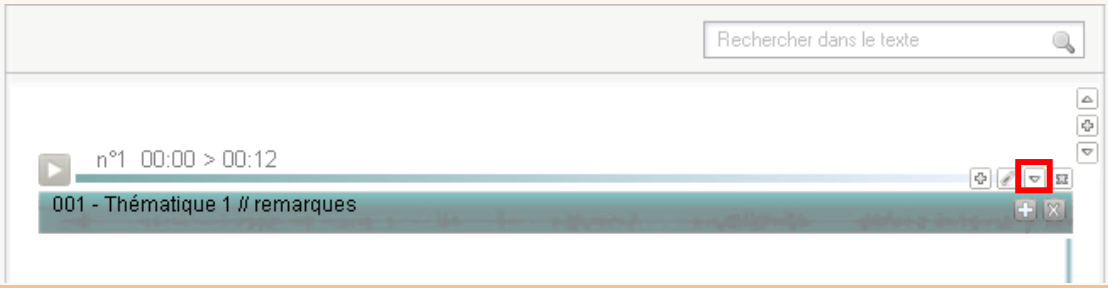
La thématisation étant une tâche de base dans Sonal, il y a de nombreux moyens d'affecter une ou plusieurs thématiques à un extrait.

Dans la zone d'édition, le plus simple est de cliquer sur le bouton ci-dessous :

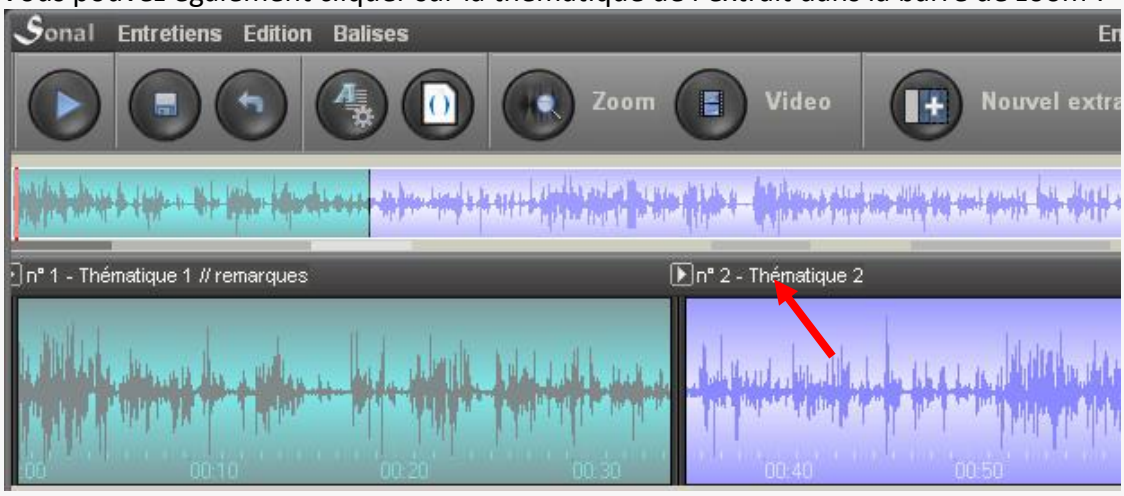


La fenêtre de choix thématique apparaît alors. Son fonctionnement est assez simple et permet la juxtaposition de plusieurs thèmes pour un extrait. Ce même bouton apparaît à tous les endroits du logiciel où il est possible de modifier la thématique des extraits.

NB : vous n'êtes pas obligés d'éditer l'extrait pour en changer la thématique. Directement dans la page de codage, vous pouvez cliquer sur le bouton ci-dessous dans l'entête d'extrait pour faire apparaître les thématiques et les modifier :

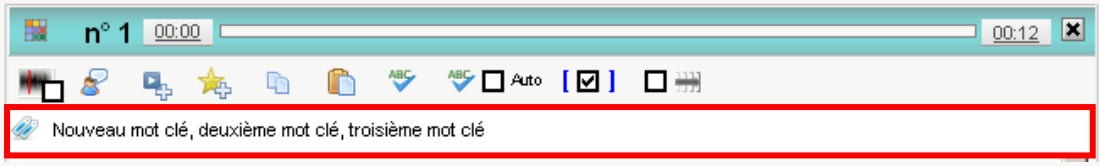


Vous pouvez également cliquer sur la thématique de l'extrait dans la barre de zoom :

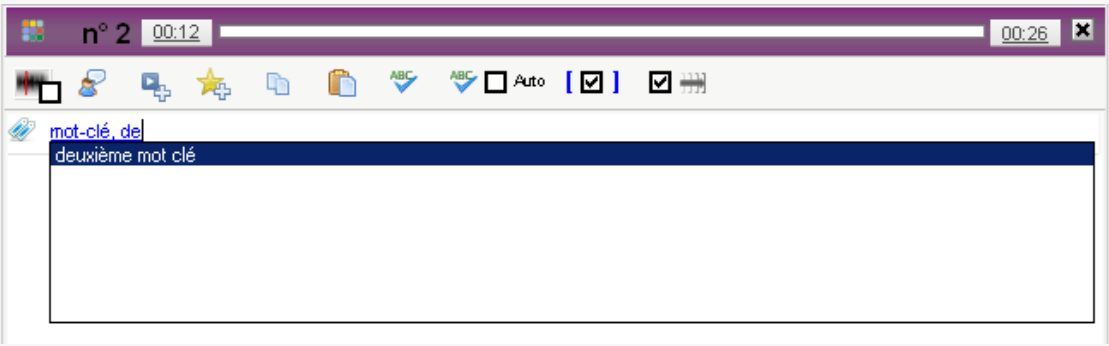


8.4 Ajouter des mots-clés (tags) aux extraits

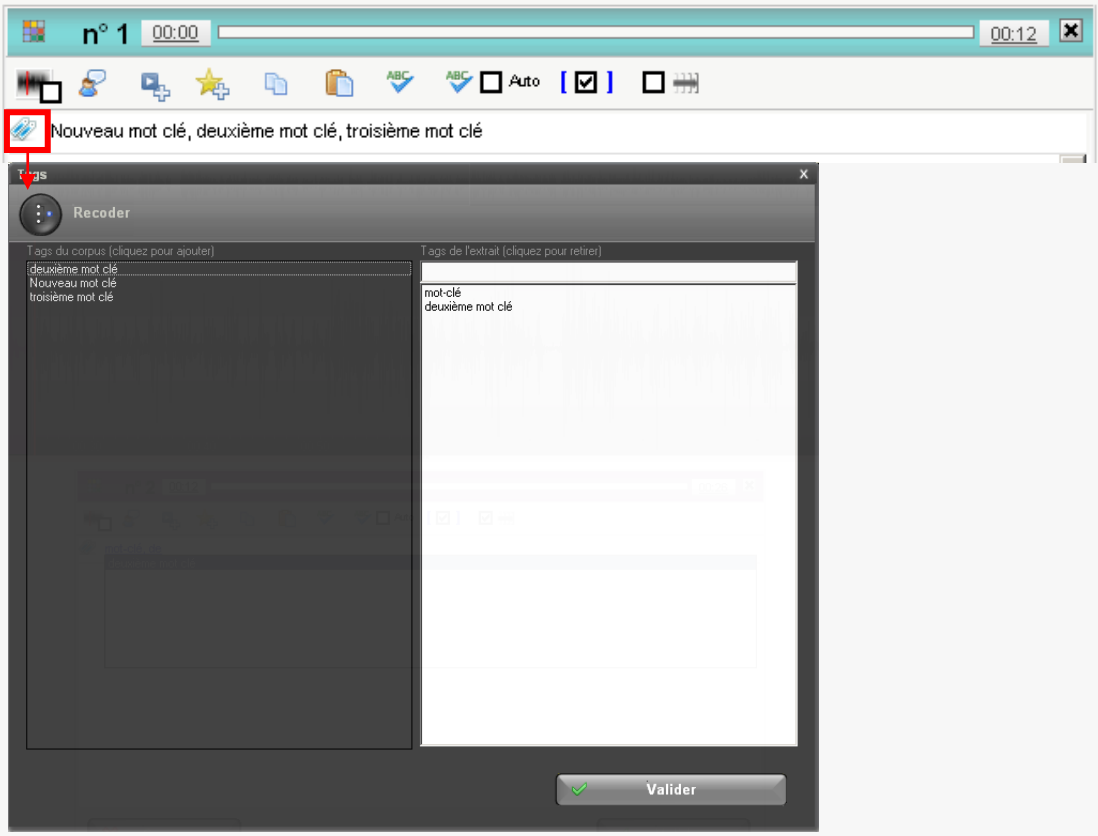
On peut ajouter des mots-clés (tags) aux extraits en les tapant dans la zone ci-dessous :



Les mots clés peuvent être des groupes nominaux, ils sont séparés par des virgules.
La saisie est semi-automatique. Si un mot-clé existe déjà dans la liste, il sera proposé ensuite à la saisie :



Vous pouvez également cliquer sur le bouton ci-dessous pour afficher la fenêtre de gestion des mots-clés (tags) :



Les mots-clés ajoutés serviront ensuite à filtrer les extraits.

8.5 Retranscrire le contenu des extraits (mode dictée)

La zone de texte (4) qui apparaît dans la fenêtre d'édition sert à affecter du texte à l'intérieur de l'extrait. Cela peut être de simples commentaires sur le contenu de l'extrait ou une retranscription intégrale.

Pour retranscrire intégralement, Sonal vous rend plusieurs services particulièrement précieux.

1) Le contrôle de la bande via le clavier (trois espaces ou alt+espace pour arrêter ou démarrer la lecture)
2) Les raccourcis de locuteurs (F1 à F12) qui mettent en forme les tours de parole

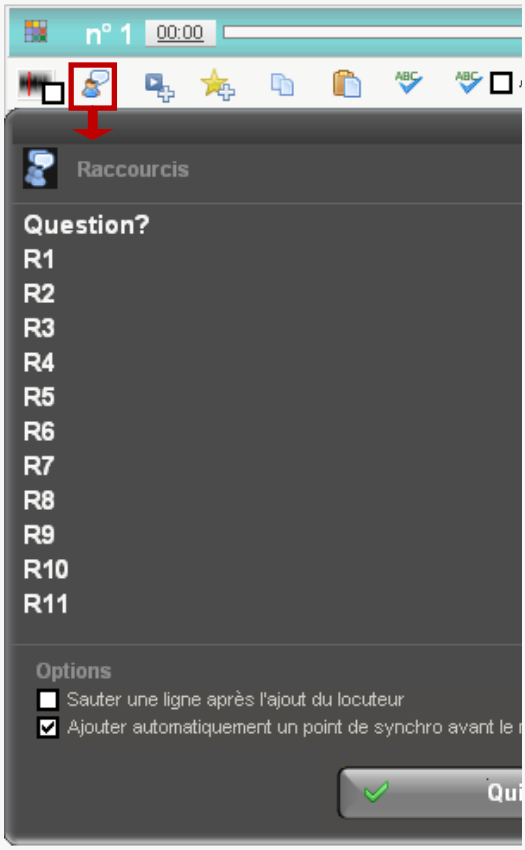
3) Le mode dictée, qui facilite grandement la retranscription.

Le contrôle de la bande au clavier ayant été vu plus haut, passons directement au paramétrage des locuteurs.

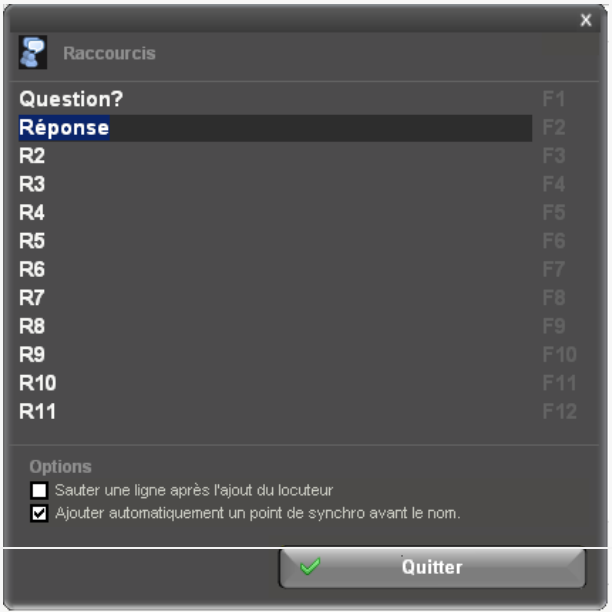
- Paramétrer les raccourcis de locuteurs

Les raccourcis de locuteurs correspondent aux touches F1 à F12. Ils permettent d'ajouter automatiquement des **balises de locuteur**, de sauter des lignes entre les interventions, et d'ajouter automatiquement des « points de synchronisation » au début des prises de parole. Les balises de locuteurs permettent le découpage du texte en interventions qui interviennent dans l'analyse quantitative des échanges ou analyse conversationnelle.

Ce que renvoient les touches F1 à F12 se paramètre directement depuis la fenêtre de retranscription, en cliquant sur le bouton ci-dessous :



Les locuteurs sont paramétrés par défaut de R1 à R11. La première ligne est réservée à la question. Pour modifier la chaine de caractères



associée par exemple à la touche F1, il faut tout simplement cliquer sur l'étiquette qui contient le code "R1", taper ce que vous voulez voir renvoyé à la place, par exemple "Question ?", puis **valider avec la touche entrée**.

En bas de page, vous pouvez utiliser deux options :

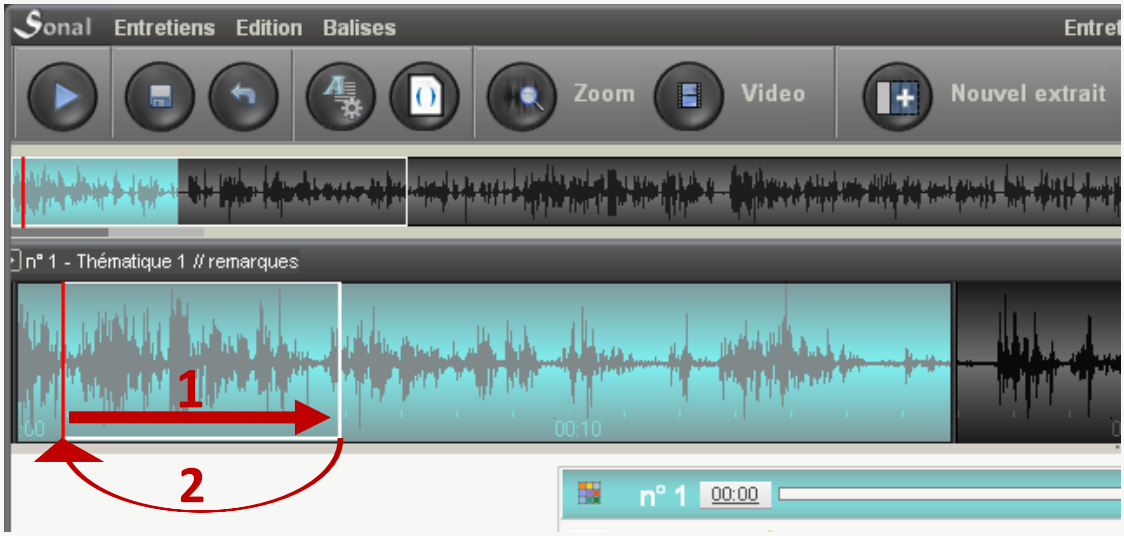
Ajouter un saut de ligne après le nom du locuteur : la fonction est explicite.

Ajouter automatiquement un point de synchro avant le nom: cette fonctionnalité vous permet de préparer le redécoupage éventuel de l'entretien en plaçant un point de synchronisation à chaque question et réponse. Vous améliorez ainsi le maillage chronologique de l'entretien, ce qui est très bénéfique par la suite. Il est fortement recommandé de laisser cette case cochée.

Une fois vos raccourcis paramétrés, vous pouvez commencer la saisie proprement dite.

- Le mode dictée

Le mode dictée est une fonction centrale de Sonal. La « dictée » contrôle la progression de la bande à votre place lors de la retranscription. Le principe général de la dictée est le suivant : lorsque le mode dictée est activé, **Sonal vous propose de retranscrire des portions de bande successives** de quelques secondes (représentées par un rectangle blanc sur la bande). Le lecteur progresse jusqu'à la fin de la portion courante, puis revient au début et s'arrête, attendant votre signal. Cela vous laisse le temps de retranscrire tranquillement avant de passer à la suite.



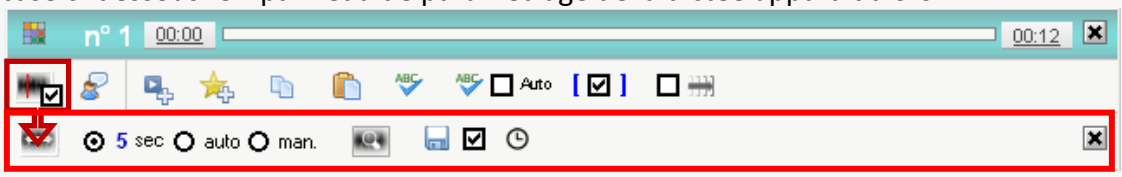
A la fin de la portion, vous pouvez la relire (alt+espace ou 2x espace) ou passer à la suivante (entrée).

En mode dictée, si vous voulez **sauter une ligne** et non pas passer à la portion suivante, **tapez alt + entrée.**

Une fois que vous passez à la portion suivant, la lecture reprend une seconde avant la fin de la portion précédente et lit la portion suivante, avant de s'arrêter. Vous pouvez de nouveau faire alt+espace pour redémarrer ou arrêter la bande, et retranscrire confortablement la portion, avant de la valider.

Tout l'intérêt de la dictée est qu'au fur et à mesure des validations, si vous débordez de l'extrait en cours, celui-ci va s'allonger progressivement, sans intervention de votre part.

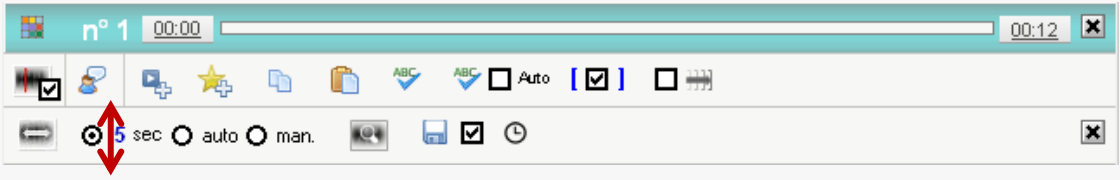
L'activation du mode dictée se réalise depuis la fenêtre retranscription en cochant la case ci-dessous. Un panneau de paramétrage de la dictée apparaît alors:



Il y a trois types de dictée différents :

- la dictée de longueur fixe,
- la dictée automatique (auto)
- la dictée manuelle (man.).

Par défaut, le mode dictée est réglé sur une durée fixe de 5 secondes.
Pour régler cette durée, passez votre souris au dessus du nombre de secondes, vous verrez un curseur de redimensionnement vertical apparaître. Cliquez dessus, maintenez enfoncé et tirez vers le haut ou vers le bas selon que vous souhaitez accroître ou diminuer la durée des portions de dictée.

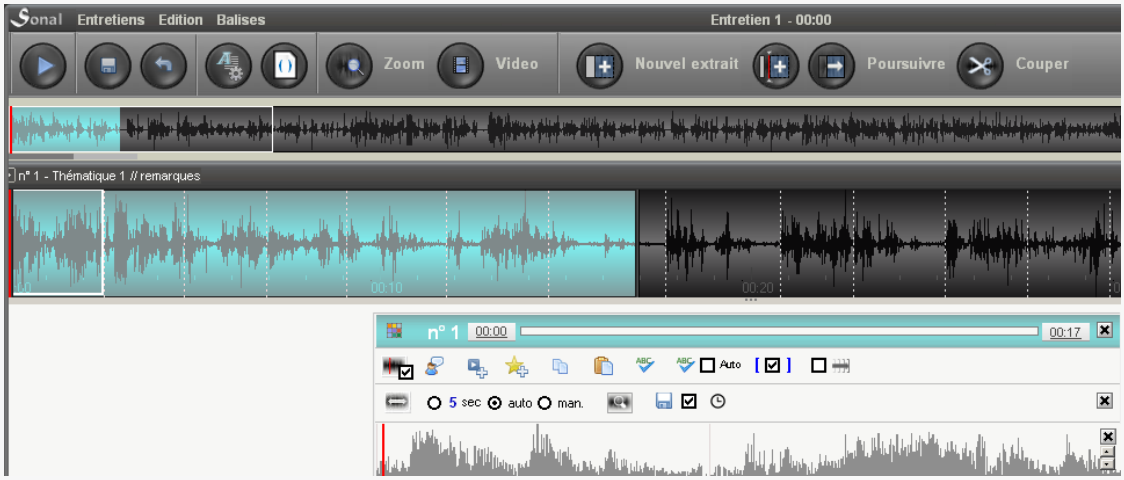


La case « auto » active la **dictée automatique.** La dictée automatique va vous permettre de **rechercher des « blancs » dans la bande de manière à mieux calibrer les portions de dictée**. Les portions ne peuvent faire moins de 2 secondes et ne dépassent pas la durée définie pour la dictée fixe. Si Sonal trouve un silence suffisamment long dans ce créneau, il le choisit comme fin de portion. Ainsi, au lieu d'avoir des passages de durée fixe qui peuvent commencer et se terminer au milieu d'un mot ou d'une phrase, la dictée automatique vous proposera des portions généralement cohérentes, ce qui simplifiera considérablement le positionnement des balises de locuteur, car il est beaucoup plus probable que les portions de dictée qui vous seront proposées correspondront peu ou prou aux alternances de parole.

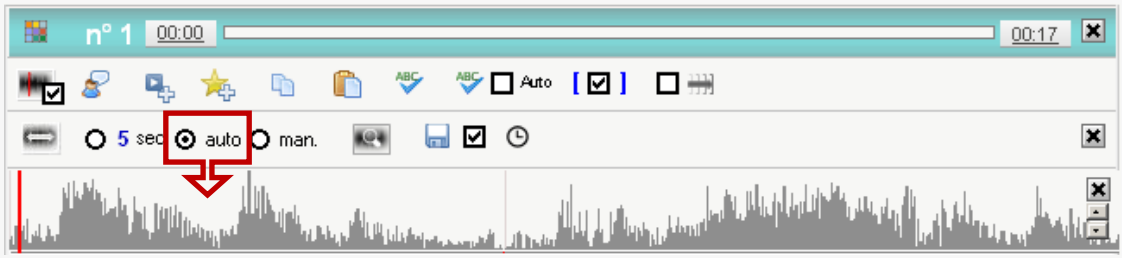
Attention, cette fonction n'est disponible qu'avec les fichiers WAV Mono et elle est beaucoup plus précise avec les fichiers encodés au format 16bits. **En phase de**

retranscription, il est donc conseillé de convertir les fichiers au format Wav, quitte à leur substituer ensuite des fichiers compressés.

Lorsque la dictée automatique est activée, Sonal commence par étudier le son de la portion considérée et propose un découpage. Cette analyse peut prendre plus ou moins de temps selon le niveau de zoom que vous avez choisi. Les portions proposées sont représentées dans la barre de zoom, à travers des lignes pointillées :



On peut voir également que lorsque la dictée automatique est activée, apparaît une « barre de calage » supplémentaire qui zoome sur le contenu de la portion de dictée :



Cette barre représente l'évolution du volume dans la portion considérée. Elle permet de voir où le logiciel propose l'arrêt dans la portion. Par exemple :

Il est possible de régler le seuil de détection du « silence ». Les boutons situés sur la droite vous permettront d'élever ou de diminuer le niveau de détection des blancs :

Attention, trop remonter le niveau de détection conduira à trouver des blancs là où il n'y en a pas, trop le baisser reviendra à n'en trouver aucun.

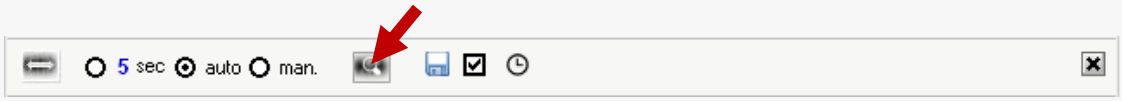
La fenêtre de dictée automatique vous permettra également de **positionner plus simplement vos points de synchronisation**. Le facteur de zoom étant élevé, il est facile de positionner précisément la position des prises de parole avec la souris. Vous pouvez alors utiliser les raccourcis locuteurs pour ajouter les balises mais il est également possible de faire un **clic droit** pour ajouter un point :

Choisissez alors le locuteur dans la liste qui apparaît : Sonal vous place un point de synchronisation suivi de la balise du locuteur en question *à la suite du texte retranscrit*.

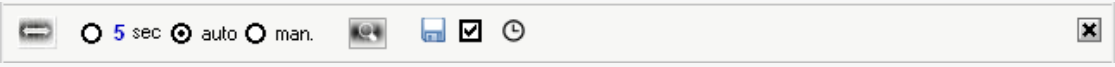
Enfin, la **dictée manuelle** (case « man. »). Il s'agit ici d'une **dictée dont les fins de portions sont définies par l'utilisateur** (via la touche entrée).
La séquence de travail est dès lors un peu différente. Lorsqu'une portion se lance, il faut **faire entrée une première fois pour arrêter son allongement**. La durée de la portion est définie. Le lecteur s'arrête et retourne au début de la portion. Prêt à être relancé. Saisissez ce que vous entendez, puis **validez à nouveau**. Cette fois-ci, la validation fera passer à la portion suivante. Validez à nouveau lorsque vous souhaitez arrêter son allongement, et ainsi de suite. Il faut faire deux fois plus souvent entrée mais au moins on a précisément les portions qui nous semblent adaptées à notre mémoire et à notre état de forme.

Le panneau de contrôle de la dictée propose ensuite différentes options :

- La vue zoomée de la portion de dictée s'affiche avec le bouton ci-dessous :



- La case en forme de disquette signifie que les modifications de l'extrait seront mémorisées à chaque validation de portion de dictée :



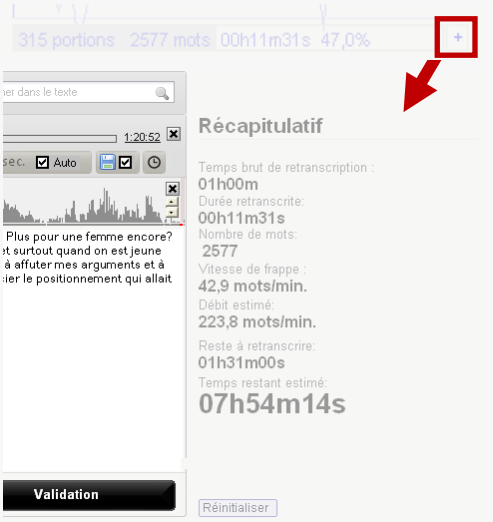
- Enfin, le bouton en forme d'horloge tout à droite sert à calculer votre temps de travail :

A chaque validation de portion de dictée, Sonal vous affiche sur le côté plusieurs informations importantes :

- Votre **vitesse de frappe** en mots par minute, avec à gauche la moyenne sur l'entretien et à droite, en gros, la vitesse de la dernière portion (verte si vous accélérez ou maintenez le rythme, rouge sinon).
- Le ratio de retranscription qui rapporte le temps de travail brut au temps de bande retranscrit. Cela vous donne toute simplement le **nombre d'heures de travail nécessaire par heure de bande** à retranscrire. Là, aussi, vous avez à gauche la moyenne et à droite le résultat pour la dernière portion de dictée.

En bas de la page, vous avez le nombre de portions retranscrites, le nombre de mots ajoutés et le temps de travail brut cumulé. Enfin, à droite, le pourcentage de l'entretien retranscrit.

Le petit plus situé tout en bas à droite vous donnera des informations complémentaires que vous pouvez ne pas vouloir connaître…



Les informations données parlent d'elles-mêmes mais quelques précisions peuvent être apportées : le temps « brut » de retranscription correspond au temps que vous passez à saisir des portions de dictée. Le temps que vous passez hors de la fenêtre d'édition des extraits n'est pas compté. Concrètement, Sonal compare la différence entre l'heure d'ouverture d'une portion de dictée et son heure de validation.

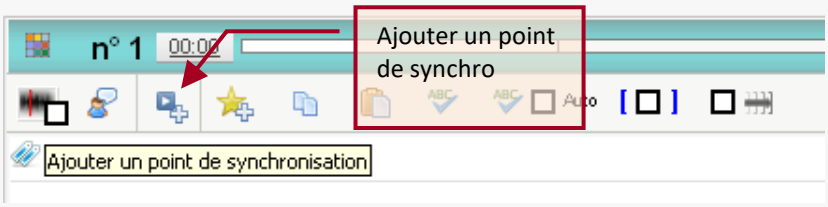
> *Le bouton « réinitialiser », en bas à gauche, vous permettra de remettre à zéro l'historique. Cela ne supprimera évidemment rien de ce que vous avez retranscrit mais vous permettra seulement de calculer vos moyennes sur des bases nouvelles.*

- Les options de la zone d'édition

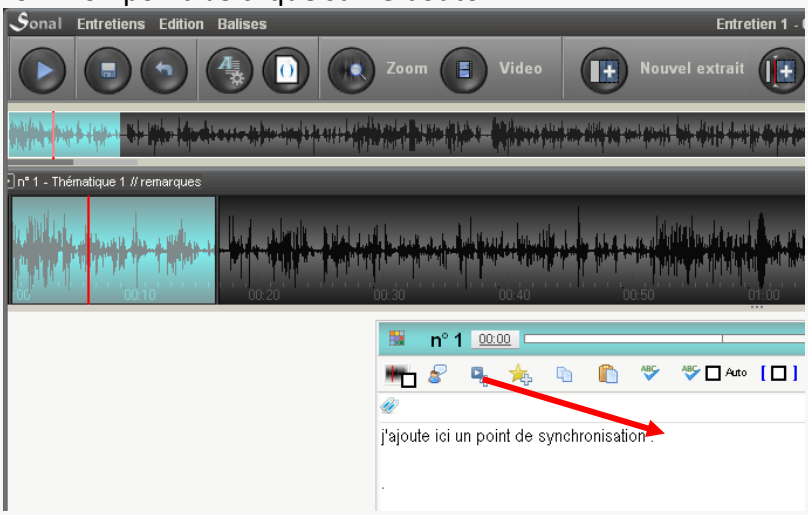
Outre la dictée et le paramétrage des locuteurs, que nous avons vu plus haut, Sonal vous propose différentes options relatives à la gestion du texte retranscrit.

- L'ajout d'un point de synchronisation

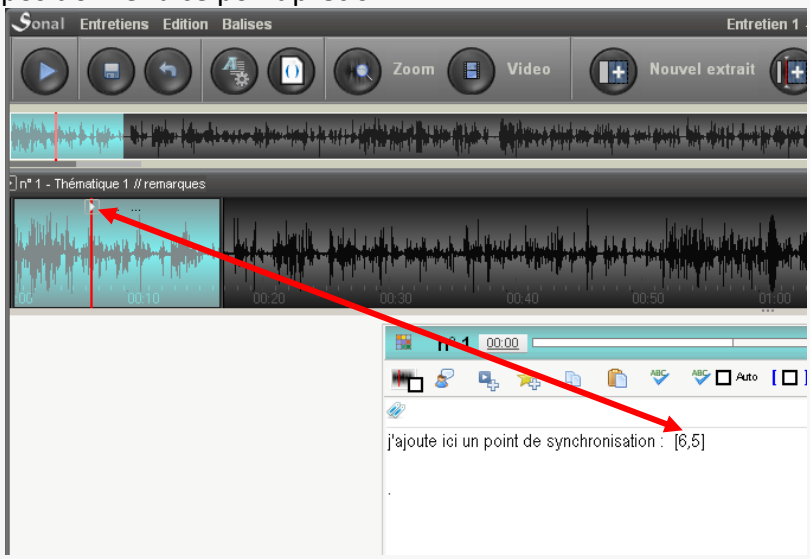
Si vous utilisez les raccourcis de locuteur et le mode dictée, vous placerez automatiquement des points de synchro un peu partout (à chaque prise de parole, à la fin des portions de dictée etc.). Vous pouvez cependant en ajouter manuellement depuis la zone d'édition (la procédure hors de cette zone est différente). Dans la zone d'édition, lorsque vous cliquez sur le bouton ci-dessous, Sonal envoie dans la zone de texte, à la position du curseur, un point de synchronisation correspondant à la position atteinte par le lecteur media.



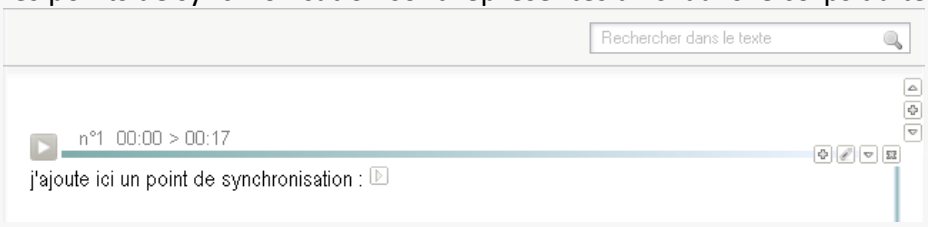
Par exemple ici, je place le curseur dans la zone de texte, à l'endroit où je veux voir mon point. Je clique sur le bouton :



Un point a été ajouté dans le texte : [6,5] et il est représenté dans la barre de zoom par un petit bouton play. Si je clique sur ce dernier, la lecture va se positionner à ce point précis.



Les points de synchronisation sont représentés ainsi dans le corps du texte :

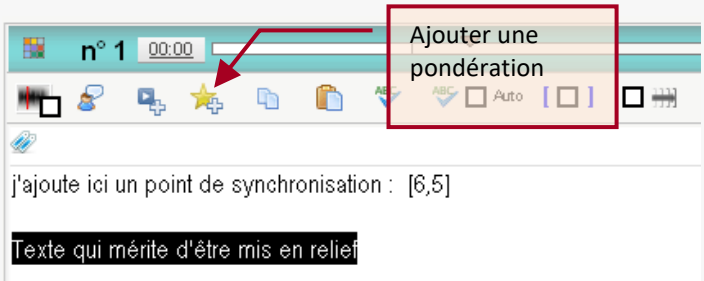


- L'ajout d'une pondération

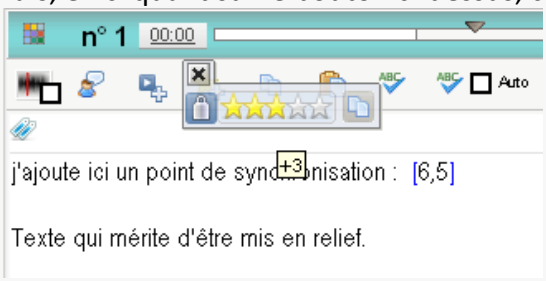
Le système de pondération qualitative de Sonal vous permet de repérer des passages dans le texte qui vous semblent intéressants en leur affectant un « poids » allant de 1 à 5. Vous pourrez ensuite isoler les parties repérées. **C'est**

particulièrement utile pour exporter des citations intéressantes. Il est particulièrement judicieux de commencer ce travail dès la retranscription.

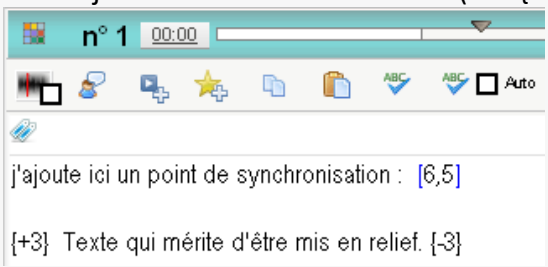
Comme pour les points de synchronisation, il y a plusieurs manières de procéder. Depuis la zone d'édition, il faut d'abord sélectionner la partie de texte que vous voulez pondérer :



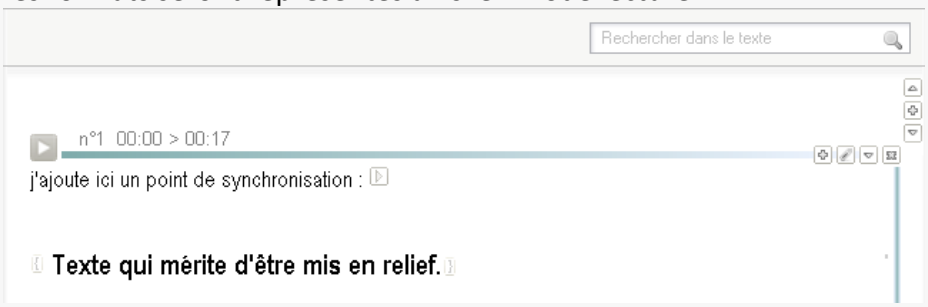
Puis, en cliquant sur le bouton ci-dessus, choisir un « poids » :



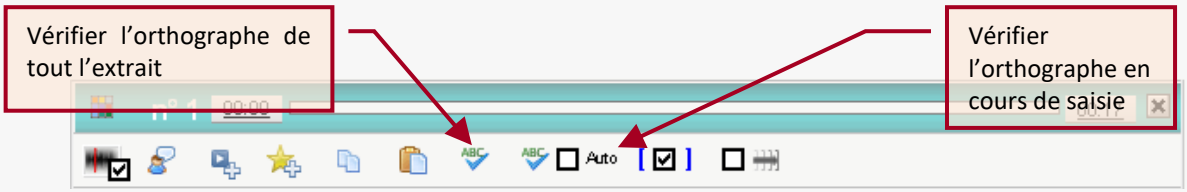
Sonal ajoute des **balises de format** (ici : {+3} et {-3}) autour du texte sélectionné.



Les formats seront représentés ainsi en mode lecture :



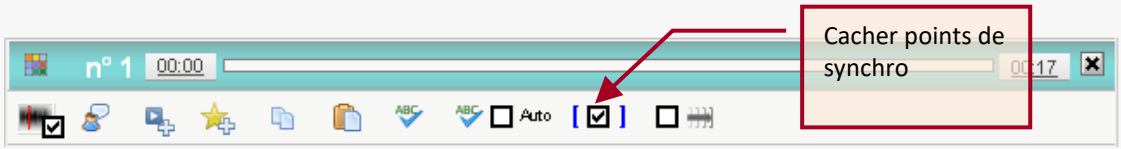
- Le vérificateur d'orthographe. Il permet de vérifier l'orthographe du texte saisi soit en cours de frappe (les mots mal orthographiés apparaissent en rouge), soit de manière globale, en vérifiant tout l'extrait.



Une chose très importante à savoir : SONAL « sous-traite » la vérification à Microsoft Word et uniquement à Word. Si Word n'est pas installé sur votre machine ou que vous avez une version non compatible, la correction ne fonctionnera pas. Des problèmes ont été mentionnés avec la version 2013 d'office…

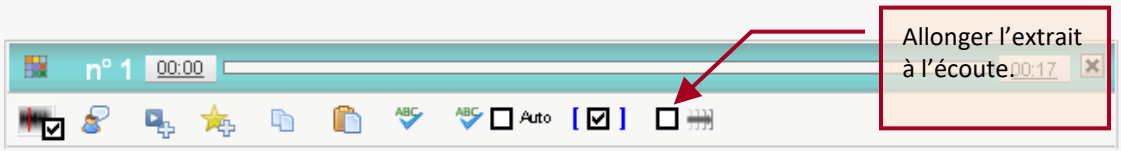
- Masquer les points de synchro dans la zone de texte éditable

La case ci-dessous permet de masquer pour partie les points de synchronisation dans la zone de texte. Lorsque cette case est cochée, les points sont symbolisés par [ ] , ce qui allège légèrement la lecture.



- Permettre l'allongement de l'extrait au fil de la lecture

La case ci-dessous permet de laisser Sonal allonger l'extrait au fil de la lecture. Si cette case est décochée, Sonal arrêtera le lecteur s'il dépasse la fin de l'extrait, sinon l'extrait s'allongera au fil de la lecture jusqu'à arrêt de la bande.



Cette case est cochée par défaut lorsque l'on crée un nouvel extrait, pour permettre un codage à l'écoute. A la réouverture des extraits, elle est décochée par défaut.

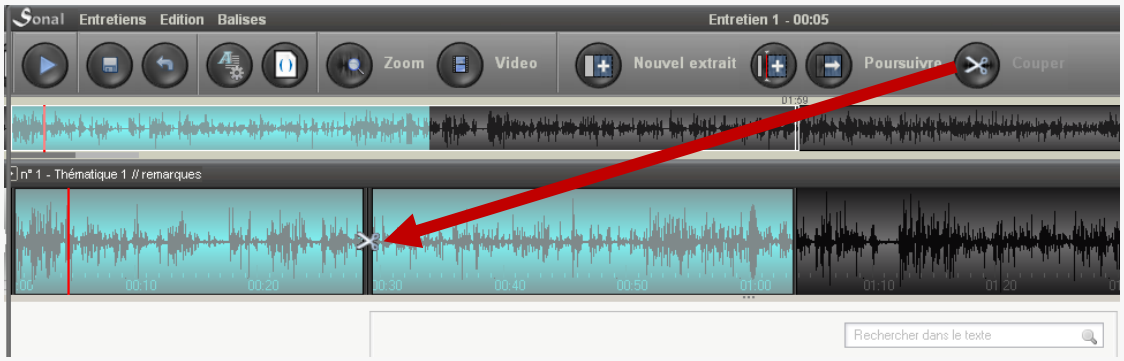
- Insérer des commentaires dans la retranscription

Vous pouvez ajouter des commentaires dans la retranscription, qui pourront être ou non affichés au moment de la lecture, et qui seront exclus de toutes les analyses (lexicométrie, dynamique des échanges etc.). **Placez les parties commentées entre parenthèses**.

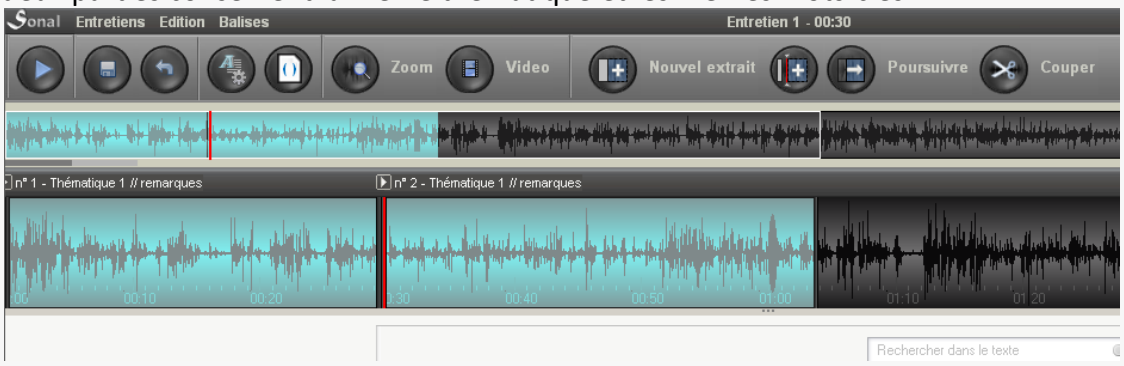
8.6 Découper un extrait

- Découper les extraits sonores

On active le mode découpage en cliquant sur le bouton « Couper ». Le bouton Couper se met alors à clignoter. Le curseur prend la forme de ciseaux lorsque l'on survole l'extrait dans la barre de zoom:



Lorsque l'on clique à un endroit sur l'extrait, celui-ci se divise en deux parties. Les deux parties conservent la même thématique et les mêmes mots-clés :



Cette méthode est simple s'il n'y a pas de texte retranscrit, mais si l'extrait contient du texte synchronisé, il vaut mieux procéder autrement. En effet, Sonal va découper le texte en se basant sur les coordonnées qu'il connaît, pour placer le texte aussi bien que possible de part et d'autre de la coupure. Mais l'efficacité de cette procédure est hautement fonction du maillage chronologique du texte.

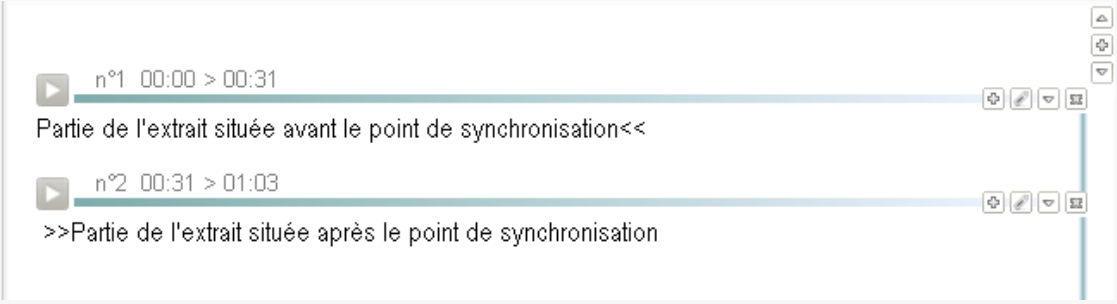
- Couper les extraits audio-textuels (à un point de synchronisation)

Le moyen le plus simple pour redécouper un extrait est de s'appuyer sur les points de synchronisation laissés dans le texte au moment de la retranscription.

En faisant un clic droit sur le point, apparaît un menu contextuel :

En cliquant sur le bouton ci-dessous, vous allez scinder l'extrait de part et d'autre du point :

Résultat, le point de synchronisation est transformé en borne d'extrait.



- Découper les extraits audio-textuels (sans point de synchronisation)

Faute d'avoir un point de synchronisation idéalement placé, il va falloir en ajouter un et l'utiliser comme bordure d'extrait. Voici comment faire.

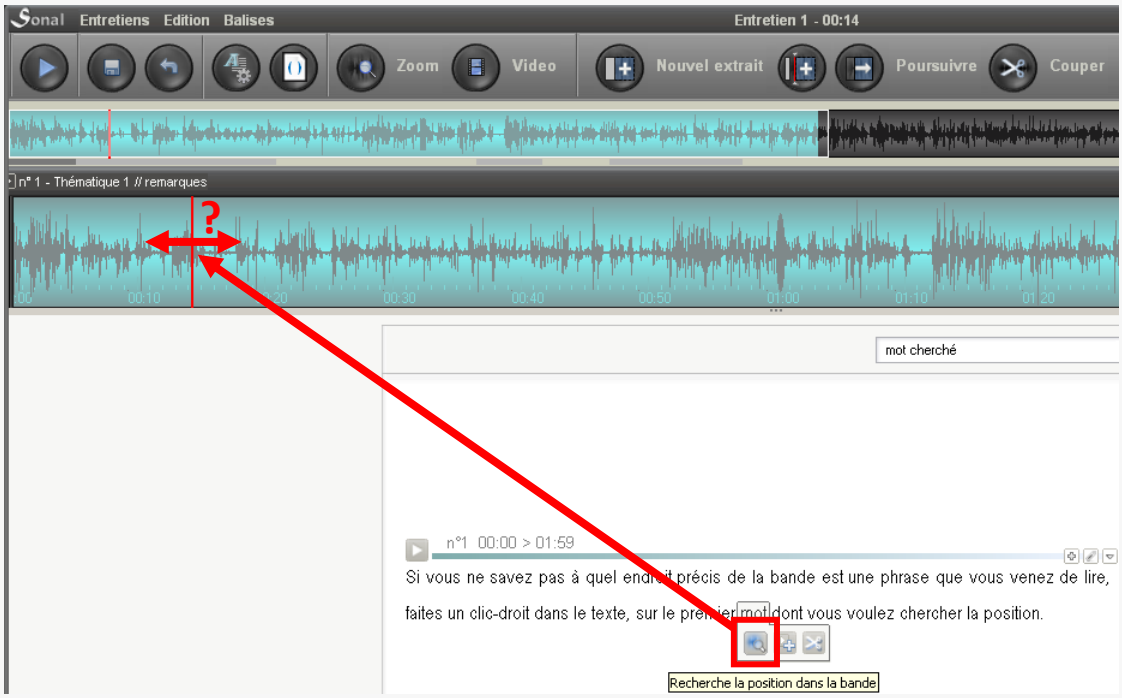
8.7 Ajouter un point de synchronisation

Pour ajouter un point de synchronisation **il faut commencer par « caler » la bande et le texte**, c'est-à-dire repérer la position (en h :m :s) dans l'enregistrement d'un mot donné ou savoir à quel mot du texte correspond une coordonnée dans l'enregistrement. Il faut reconstituer la synchronisation. Heureusement, Sonal vous aide un peu. Selon que vous souhaitez partir du texte ou du son, voici comment procéder.

- Trouver la position d'un mot dans l'enregistrement:

Si vous ne savez pas à quel endroit précis de la bande est une phrase que vous venez de lire, faites un clic-droit dans le texte, sur le premier mot dont vous voulez chercher la position. Dans le menu contextuel qui apparaît, cliquez sur le bouton ci-dessous :

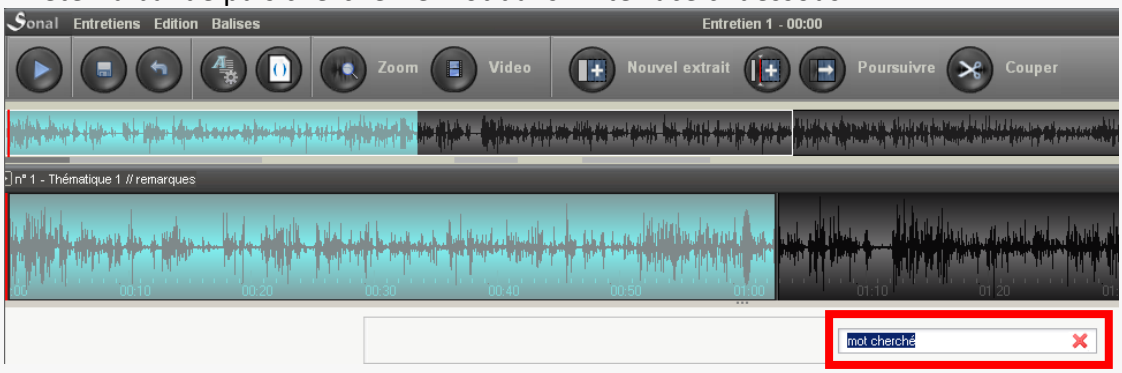
Sonal vous proposera alors, **sur la base d'un calcul**, une position supposée dans la bande :



Selon la qualité du maillage chronologique et la qualité de retranscription de votre entretien, le calage sera plus ou moins précis. Vous pouvez vous retrouver quelques secondes avant ou après la vraie position du mot. Reste que Sonal vous aura dégrossi un peu le travail. En écoutant la bande et en lisant les passages alentours, vous allez trouver votre mot. Au besoin, si vraiment vous êtes perdus, alternez avec la méthode ci-dessous :

- Trouver la position dans le texte d'un passage sonore :

Repérez un mot peu fréquent dans le son proche du point que vous voulez placer. Arrêtez la bande puis cherchez le mot dans l'interface ci-dessous :

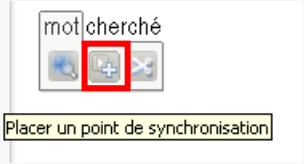


Faites « entrée » pour lancer la recherche dans le texte. Une fois le bon mot trouvé dans le texte, repositionnez au besoin le lecteur sonore pour assurer la correspondance entre son et texte.

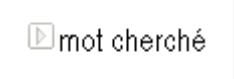
Vous pouvez alors ajouter soit un point de synchronisation, soit un « point de partition », c'est-à-dire une bordure d'extrait. Dans un cas vous vous contentez de mémoriser la position, dans l'autre vous utilisez cette position pour découper votre extrait en deux. Dans tous les cas, vous pourrez à loisir transformer un point de synchro en point de partition et réciproquement.

- Placer un point de synchronisation

Une fois que vous avez « calé » le texte et le son, faites un clic-droit sur le mot avant lequel vous voulez placer le point de synchro et, dans le menu contextuel qui apparaît, cliquez sur le bouton suivant :

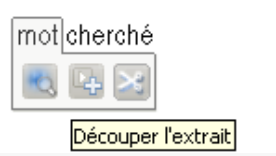


Sonal placera alors un point de synchro devant le mot voulu :



- Placer une bordure d'extrait (point de partition)

Une fois que vous avez « calé » le texte et le son, faites un clic-droit sur le premier mot de l'extrait à créer puis, dans le menu contextuel qui apparaît, cliquez sur le bouton suivant :



Sonal coupera alors l'extrait de part et d'autre de ce mot, en créant deux extraits audio-textuels cohérents. Exactement comme si vous aviez ajouté un point de synchronisation avant de l'utiliser pour couper votre extrait (voir ci-dessus).

8.8 Corriger un point de synchronisation

Comme les points de synchronisation sont laissés dans le texte, vous pouvez évidemment éditer vos extraits et les modifier à la main directement dans le texte, mais Sonal vous propose une méthode plus simple.

- Corriger un point manuellement

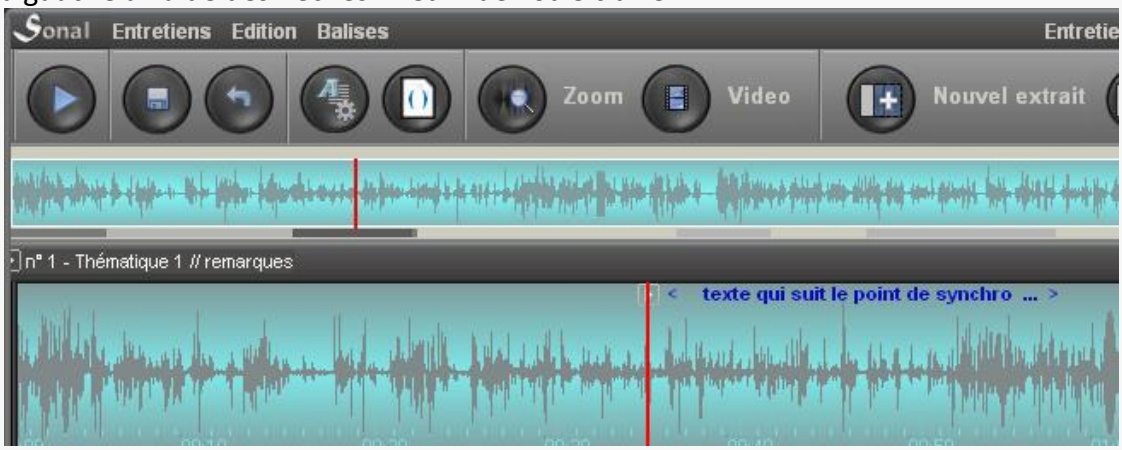
Vous pouvez corriger les points par glisser/déplacer dans la barre de zoom. Pour ce faire, il faut d'abord cliquer sur le **texte qui suit le point**.

Le point change alors d'apparence, il est déplaçable à droite et à gauche :

Une fois le point relâché, il reprend sa forme initiale.

- Corriger un point au clavier

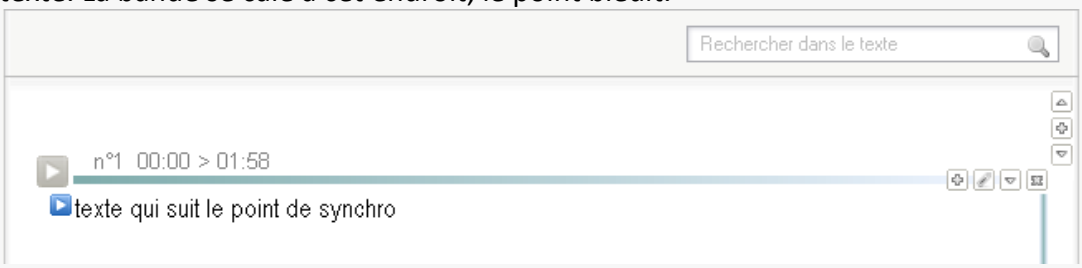
Vous pouvez passer de point en point en utilisant les flèche <- et -> de votre clavier. **Quand un point est sélectionné, il est mis en gras**. Tapez alors « entrée » sur votre clavier, le texte du point devient bleu et peut être déplacé d'une seconde à droite et à gauche à l'aide des flèches <- et -> de votre clavier :



Une fois votre point calé, faites entrée. Il redevient normal et vous pouvez passer au suivant.

- Déplacer un point dans le texte

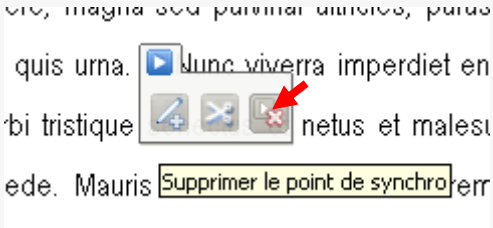
Vous pouvez aussi choisir non pas de corriger la position chronométrique de votre point mais, en le laissant à sa position chronométrique, le placer au bon endroit *dans le texte*.
Pour ce faire, cliquez une première fois sur le point que vous voulez déplacer dans le texte. La bande se cale à cet endroit, le point bleuit.



Après ce premier clic, le point peut être déplacé par glisser-déposer. Déplacez-le sur le mot devant lequel vous voulez le placer. Relâchez. Il est repositionné.

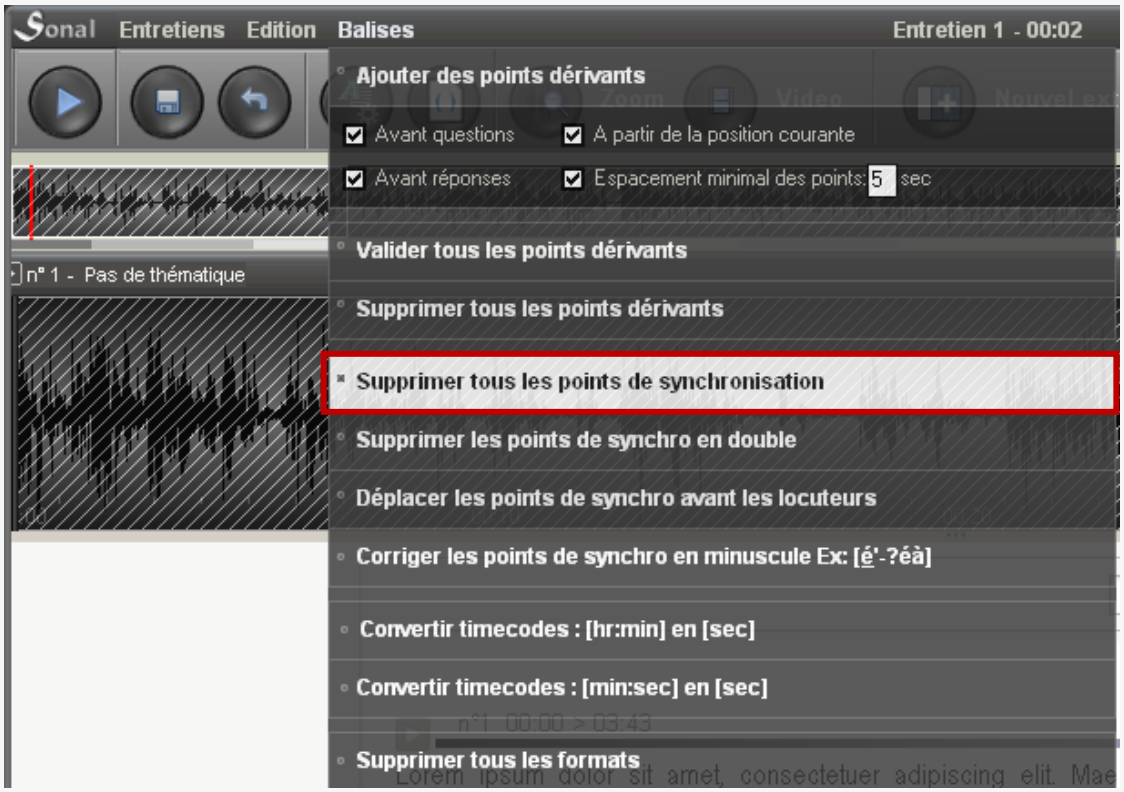
8.9 Supprimer un point de synchronisation

Au clic droit, les points de synchro (dans la barre de zoom comme dans le texte) font apparaître un menu contextuel qui vous permet de supprimer les points en cliquant sur le bouton ci-dessous :



8.10 Supprimer tous les points de synchronisation

Il est parfois utile de supprimer tous les points de synchronisation d'un entretien. Pour ce faire, on peut cliquer sur le menu « balises » de la fenêtre de codage, puis sur le menu suivant :



Juste en dessous, vous avez également la possibilité de supprimer les points en double.

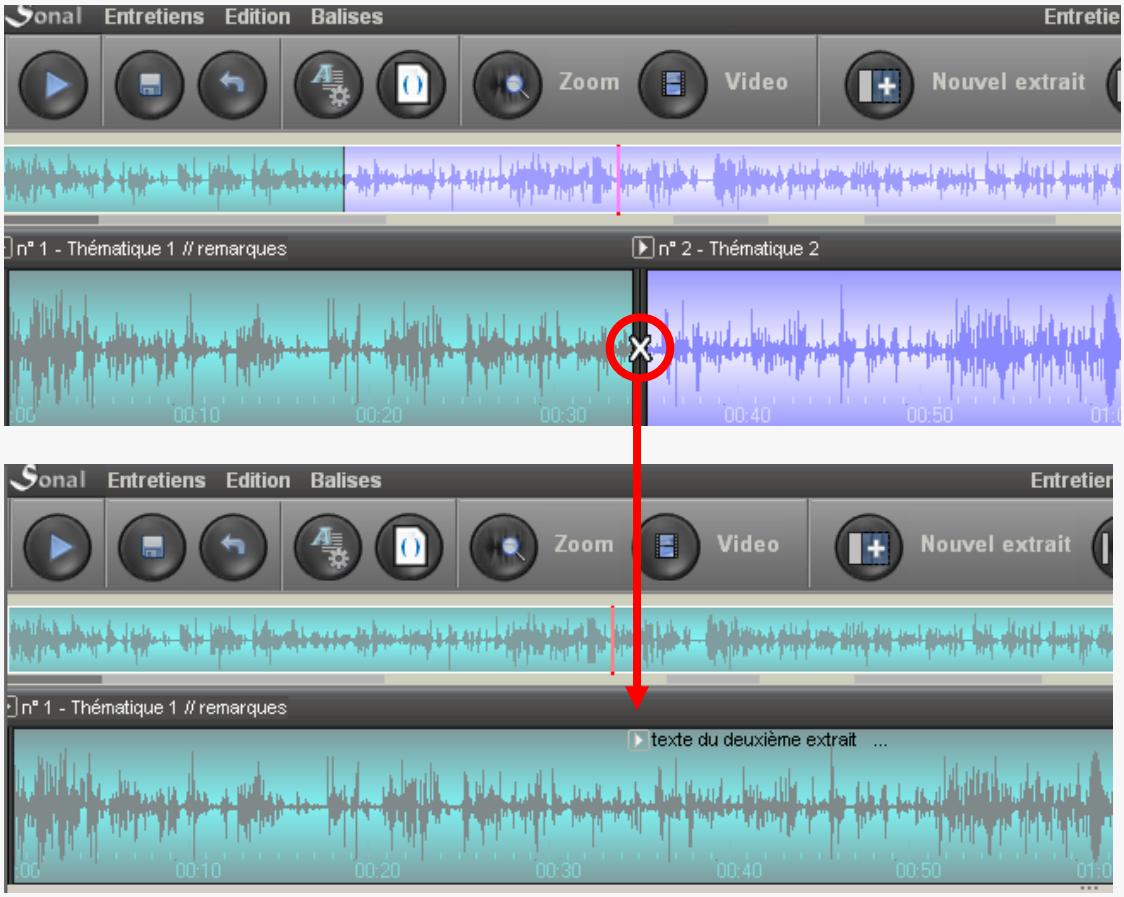
8.11 Fusionner deux extraits

La fusion de deux extraits (ex : extraits A et B qui se suivent) revient à ne créer qu'un extrait à partir de deux extraits *contigus*. La fusion a les conséquences suivantes :

- Le nouvel extrait créé débute au début de l'extrait A et termine à la fin du B
- Le nouvel extrait créé conserve la thématique de l'extrait A
- Le texte des deux extraits est concaténé de part et d'autre d'un point de synchro qui mémorise la position de l'ancienne coupure (A + point de synchro + B)
- Les mots-clés des deux extraits s'ajoutent (attention aux doublons !)

Il y a plusieurs manières de fusionner deux extraits :

Si les deux extraits sont contigus, vous pouvez cliquer sur la jonction des deux extraits dans la barre de zoom.

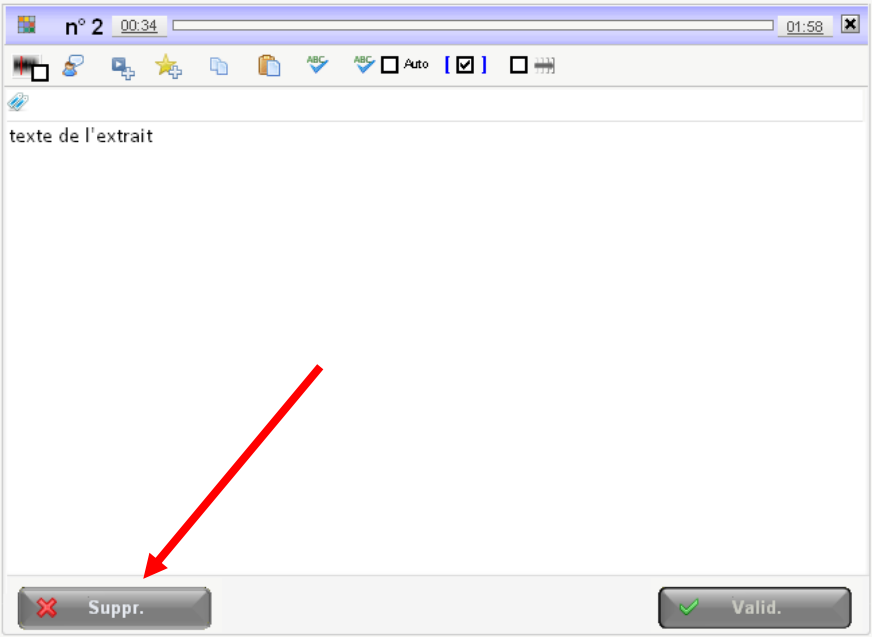


Vous pouvez aussi cliquer sur la petite croix tout à droite de la ligne de titre de l'extrait :

NB : cette croix veut bien dire *supprimer la bordure d'extrait* et non supprimer l'extrait. C'est la raison pour laquelle lorsque l'on clique sur celle du premier extrait, Sonal prévient que c'est impossible, car il n'y a pas d'extrait avant le premier.

8.12 Supprimer un extrait

Pour supprimer un extrait (et non le fusionner avec celui qui le précède ou lui succède), éditez le et cliquez sur le bouton « supprimer » en bas à gauche de la zone d'édition. C'est le seul endroit pour le faire.



8.13 Sauvegarder le codage

Il faut savoir que lorsque vous utilisez le mode dictée, **le texte que vous saisissez est sauvegardé à chaque validation de portion (entrée)**. Vous pouvez également taper ctrl+S depuis l'interface de saisie.

De même, les modifications de l'entretien sont sauvegardées à chaque validation d'extrait.

Au besoin, vous pouvez aussi demander l'enregistrement de la retranscription en cliquant sur le bouton ci-dessous.

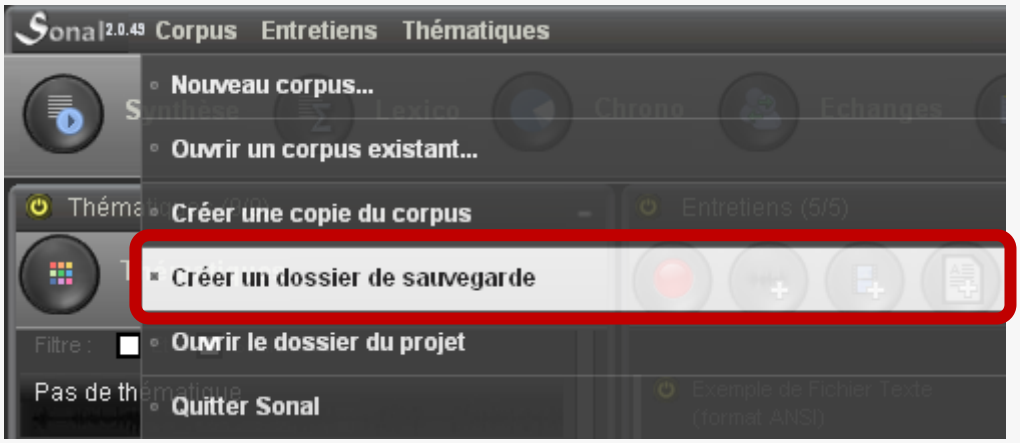
- Que se passe-t-il lorsqu'on enregistre ?

L'enregistrement revient à réécrire le fichier *.Rtr de l'entretien avec les informations courantes. **Ce fichier contient tout votre travail** sur l'entretien, il porte le même nom que la bande (voir détails sur le système de fichiers).

En parallèle, à chaque sauvegarde, une copie de ce fichier est créée dans le dossier « Restaur » situé dans le répertoire de projet. 10 versions successives sont sauvegardées, numérotées entre parenthèse (1) à (10). Pour restaurer ces versions précédentes depuis Sonal, utilisez le bouton ci-dessous :

En cas de difficulté, il est toujours possible d'aller copier la version la plus récente de ce fichier dans le dossier Restaur, et de le coller à la place du Rtr courant, en retirant son numéro à la fin du nom.

Pour parer à toute éventualité, vous pouvez réaliser un dossier de sauvegarde, depuis la fenêtre de corpus :

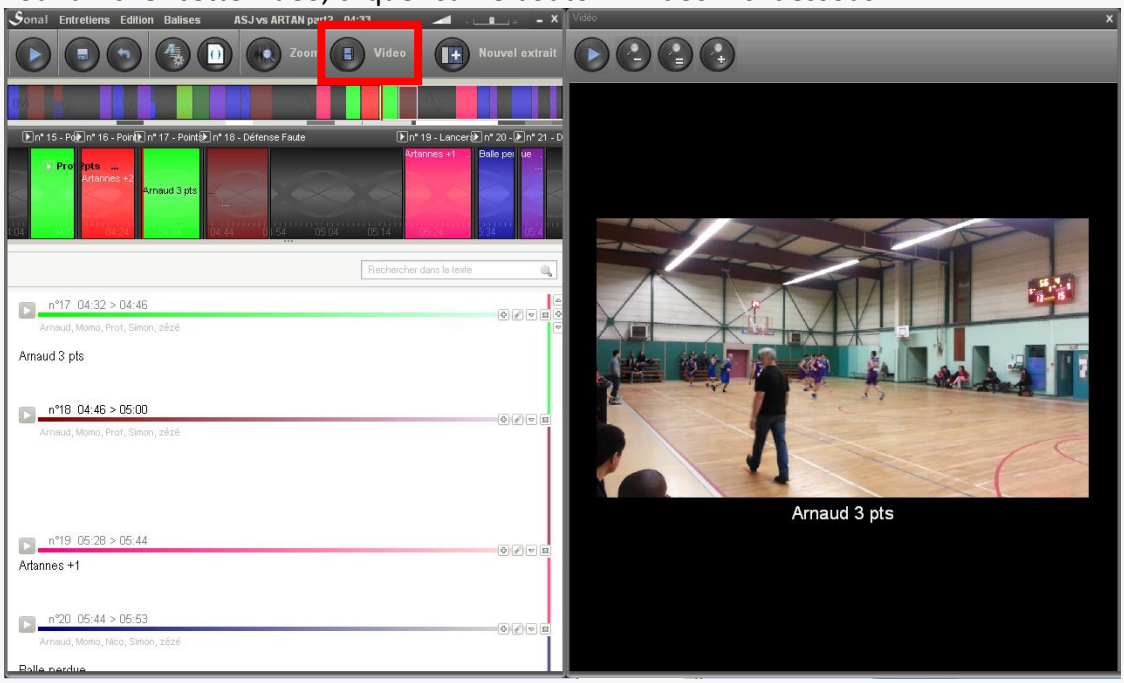


Sonal créera, au nom demandé, un dossier qui contiendra une copie des fichiers Rtr de tous les entretiens du projet, et du fichier de corpus. Ces fichiers seront ensuite à votre disposition pour une restauration en cas de problème. Les fichiers en question sont très légers, vous ne risquez pas la surcharge en créant fréquemment des dossiers. **Faites-le à chaque fois que vous aurez bien avancé dans votre travail**.

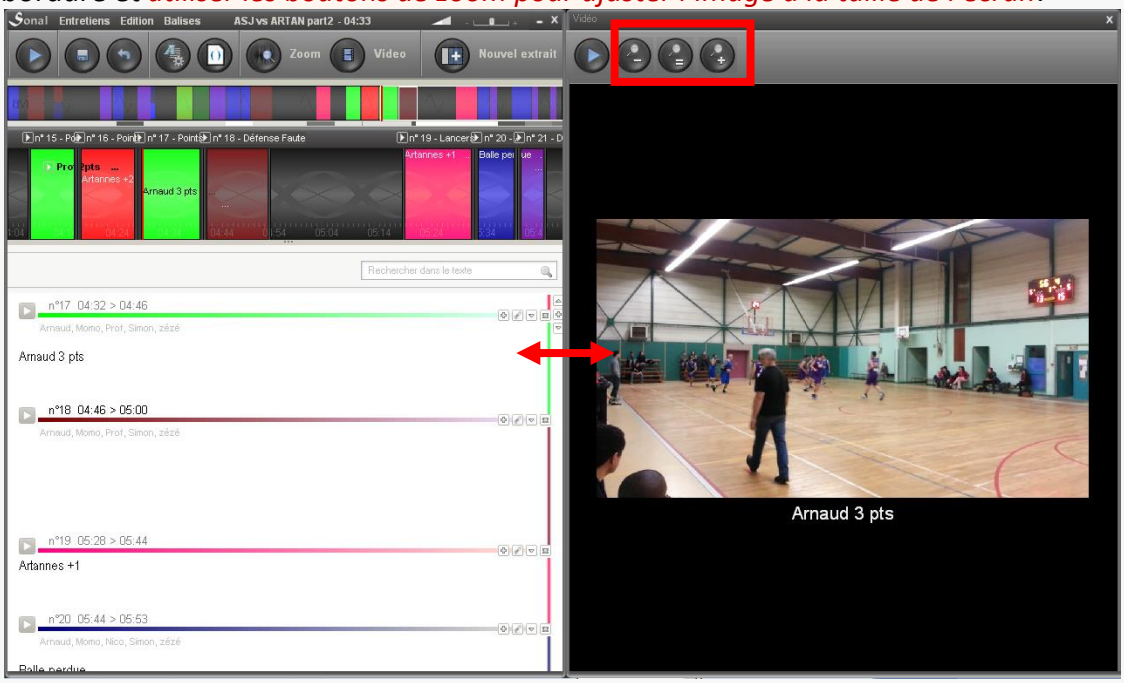
8.14 Travailler avec de la vidéo

Tout ce qui a été présenté ci-dessus s'agissant du codage vaut également pour la vidéo. La seule différence est que vous aurez la vidéo à droite de la page de codage.

Pour afficher cette vidéo, cliquez sur le bouton « vidéo » ci-dessous :



Le panneau latéral apparaît alors avec la vidéo. Vous pouvez l'élargir en tirant sur la bordure et *utiliser les boutons de zoom pour ajuster l'image à la taille de l'écran*.



On peut également remarquer que les parties du texte qui sont repérées par leur position (débuts d'extrait, points de synchro) sont utilisés pour sous-titrer la vidéo. Autrement dit, **il est tout à fait possible d'utiliser Sonal comme outil de sous-titrage**. Un export au format *.Srt est en préparation.

9 Resynchroniser une retranscription

Il arrive que l'on dispose d'un côté d'un enregistrement et de l'autre de la retranscription et que l'on souhaite faire en sorte de pouvoir resynchroniser les deux, c'est-à-dire ajouter dans le texte des marqueurs chronométriques (point de synchro), voire segmenter solidairement les deux matériaux pour recréer des extraits audio-textuels.

- Dans quels cas resynchroniser ?

La resynchronisation peut vous permettre, si vous découvrez Sonal au milieu d'une recherche, d'intégrer des entretiens déjà retranscrits dans votre corpus, pour pouvoir les soumettre au même cadre analytique que ceux que vous saisirez dans Sonal. Cela ne présente cependant d'intérêt que si vous tenez particulièrement à garder la dimension chronométrique dans vos analyses (durées, débits de parole etc.) ou si vous voulez sous-titrer une vidéo. Si seules fonctions de synthèse et d'analyse lexicométrique vous intéressent, vous pouvez aussi vous contenter de découper le texte, sans lui associer de bande (voir le partie encoder les documents textuels). Cela vous prendra moins de temps qu'une resynchronisation, qui demande quelques opérations préalables.

9.1 Etape 1 : Importer la version sonore de l'entretien

Pour resynchroniser un entretien, il vous faut d'abord avoir sa version sonore dans le corpus. Pour cela, reportez-vous à la partie consacrée à l'ajout de documents sonores.

9.2 Etape 2 : Importer le texte de la retranscription

Pour associer une retranscription à la bande, ouvrez le document que vous venez d'importer avec la page de codage/lecture de l'entretien.

Ensuite, dans le menu "Entretiens", choisissez, en fonction du format du fichier qui contient la retranscription de l'entretien l'un ou l'autre des items du menu. Vous avez deux formats possibles. Si votre fichier est au format word, choisissez "doc ou docx".

Si vous n'avez pas Word, exportez votre fichier dans le bloc note au format *.txt et choisissez ce format pour l'importation.

Ciblez le fichier qui contient les retranscriptions, choisissez "ouvrir".

Attention, si vous aviez déjà commencé à encoder le fichier, tout ce que vous avez fait sera écrasé! Un message vous prévient au cas où. Assurez-vous de ne pas faire de fausse manipulation.

Après importation, le texte du fichier va être associé à l'entretien dans un **seul et unique extrait dénué de thématique**, dont la durée sera égale à celle de l'entretien. La barre de navigation en haut de la page se strie de gris (ce qui renvoie à l'absence de thématique) et le texte de l'entretien apparaît dans son intégralité :

En l'état, il ne sert pas à grand chose d'avoir un extrait gigantesque. Il va falloir le redécouper progressivement. Mais avant, il pourra être très utile de reformater les balises, pour permettre à Sonal de repérer le changement de locuteur et le statut interrogatif des questions.

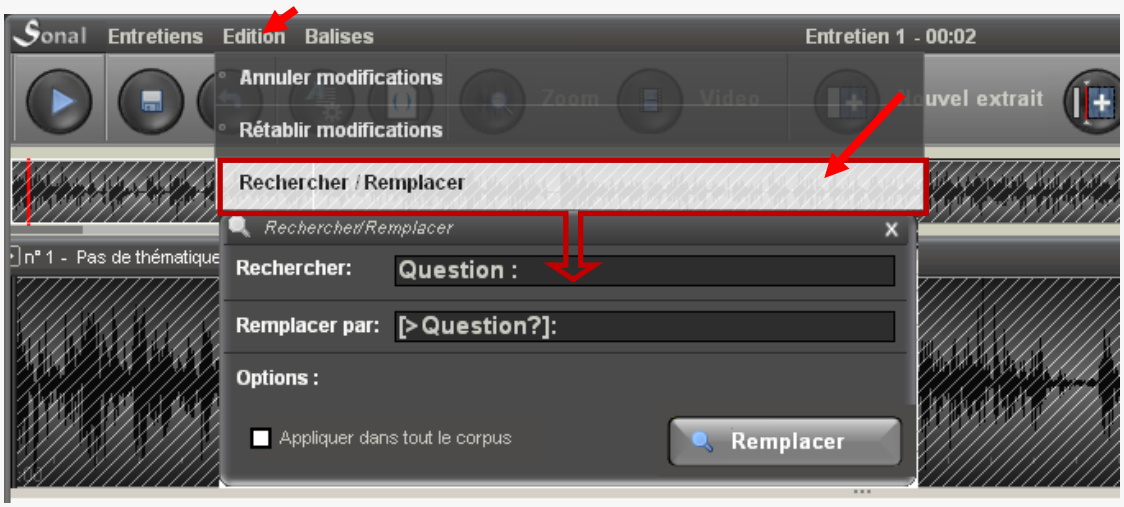
9.3 Etape 3 : Reformater les balises de locuteurs

Les entretiens retranscrits comportent généralement des indications sur les tours de parole, au minimum Question : / Réponse :

Le reformatage des balises aboutira à ce que les "Question:" soient convertis en " [>Question ?] :" afin que Sonal puisse les interpréter correctement.

> *Cela présente un avantage également si vous souhaitez utiliser la resynchronisation semi-automatique (voir plus bas). Celle-ci vous propose en effet une synchronisation appuyée sur les alternances des questions et des réponses. Si Sonal ne trouve pas de balises de locuteurs, il ne pourra pas le faire.*

Pour ce faire, une fonction rechercher/remplacer est disponible dans la page de lecture.

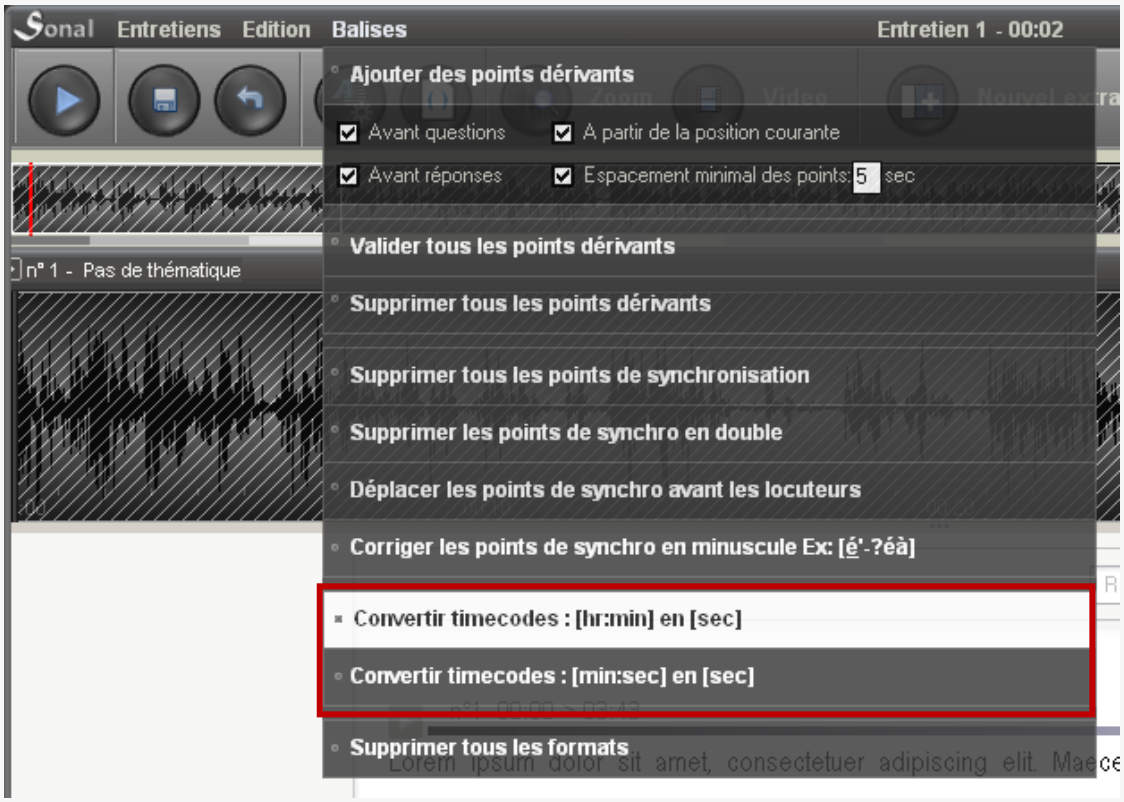


Attention, cette fonction rencontre parfois des difficultés avec les caractères importés de Word, qui ne sont pas toujours reconnus. Une méthode simple et efficace consiste à faire un rechercher/remplacer directement dans le fichier txt avant importation.

9.4 Convertir les timecodes en points de synchro (éventuellement)

Il est possible que votre retranscription contienne des timecodes ou timestamps. Ces timecodes peuvent être soit au format [min:sec] soit au format [hr:min].

Le menu ci-dessous vous permet de convertir ces timecodes en points de synchronisation lisibles par Sonal (cela revient à convertir les balises en secondes) :



Il est important que les timecodes soient les plus précis possibles. Mentionner la minute sans préciser les secondes aboutit à une erreur de placement qui peut logiquement aller jusqu'à… une minute. C'est beaucoup, beaucoup trop pour Sonal. Il faut donc que les coordonnées dans le texte correspondent bien à celles dans la bande. Il **vaut mieux ne pas avoir du tout de points de synchro que des points de synchro erronés**.

Si, après conversion, vous vous apercevez que les times codes sont trop imprécis et que les points de synchronisation sont mal placés, une fonction vous permet de débarrasser les entretiens de tous leurs points de synchronisation. Elle est accessible dans le menu "balises" de la page de lecture (voir plus haut).

9.5 Etape 4 : premier découpage grossier

Quelle que soit la méthode de resynchronisation choisie, il est précieux de commencer par un découpage en quelques grands extraits de 15 à 30 minutes. **Il vaut mieux avoir plusieurs extraits relativement courts, dont les positions sont fiables, plutôt qu'un seul très long extrait**. L'inférence des positions, lors de la resynchronisation n'en sera que meilleure, et tout le logiciel travaillera plus rapidement.

> *Le choix des points de coupure n'a pas une grande importance. L'endroit importe peu car cette coupure pourra être convertie par la suite en un simple point de synchronisation, ou même être effacée, lors de l'encodage final. Vous pouvez donc placer des coupures au milieu d'interventions longues, même s'il*

> *est préférable, pour gagner du temps, de les placer d'emblée sur des prises de parole.*

Important : il y a parfois des éléments non retranscrits en fin de bande, par exemple, quelques minutes d'échanges informels qui sont présents sur la bande mais pas dans le texte. Cela peut être problématique pour l'estimation des positions, pour des raisons assez logiques. Ecouter immédiatement la fin de l'entretien et vérifier que la bande et le texte s'arrêtent bien ensemble peut vous faire gagner du temps.

Pour savoir comment découper de premiers grands extraits, reportez vous à la partie découper des extraits.

9.6 La synchronisation semi-automatique

Il existe un moyen d'accélérer le travail de resynchronisation, sans qu'il y ait à ajouter manuellement tous les points. Comment? Par la **création semi-automatique de** points de synchronisation à chaque changement de locuteur.

> *D'où l'intérêt d'avoir converti les "question:" et "réponse:" en balises de locuteurs lisibles par Sonal (voir plus haut).*

Ces points sont dits « dérivants ». Ils s'agit d'une forme particulière de point de synchronisation.

- Ajouter des points dérivants

On ajoute des points dérivants en cliquant sur le menu suivant :

Avant de cliquer sur « Ajouter des points dérivants », choisissez les options immédiatement sous le bouton : où les points doivent être placés, à quelle fréquence etc.

Des points dérivants vont apparaître là où vous l'avez demandé.

> *Là aussi, le calcul est d'autant plus long que les extraits le sont eux-mêmes. Le prédécoupage de l'entretien en quelques grands extraits accélèrera donc le processus.*

- Corriger les points dérivants

Il faudra ensuite avancer dans l'entretien en validant successivement ces points. Pour valider un point dont la position est correcte, il suffit de cliquer dessus. Sinon, décalez-le à la souris jusqu'à sa position réelle.

La correction de plusieurs centaines de points peut prendre un certain temps. Pour accélérer ce travail, des raccourcis sont disponibles.

Les touches <- et -> du clavier font passer de point en point, et lorsque l'on valide avec entrée, le point en cours devient déplaçable avec ces mêmes touches (+ ou – 1 sec).

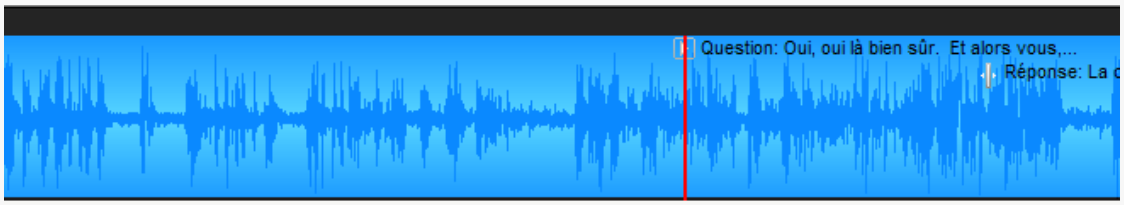
Arrivée sur un point dérivant (flèche droite):

Validation, avec entrée. Le point devient sensible au déplacement (il est bleu et des caractères < et > sont placés à ses extrémités). Si vous utilisez les flèches <- et ->, le point se déplacera vers la gauche ou vers la droite.

Astuce : si le point a une position trop anticipée (trop à gauche), pour éviter d'avoir à le décaler un grand nombre de fois vers la droite, vous pouvez taper **ctrl+entrée, le point est alors validé à la position atteinte dans la lecture**.

Une fois la position ajustée, vous pouvez revalider

Après validation, le point redevient normal et les flèches retrouvent leur fonction originale. On passe ainsi de point en point et c'est un réel gain de temps.



Le raccourci ctrl+x permet de convertir les points de synchronisation en points de partition (couper les extraits de part et d'autre).

> *Avec cette méthode, et dans de bonnes conditions, on peut resynchroniser les entretiens dans un temps compris entre un tiers et la moitié de la bande elle-même, **soit 20 à 30 minutes pour une heure d'entretien**, mais cela dépend de la qualité de la transcription et du nombre de locuteurs. Cela dépend également de la fréquence d'ajout des points dérivants. Vous n'êtes pas obligés de placer trop de points dérivants ni de tous les corriger. Vous pouvez vous focaliser sur ceux qui correspondent à un changement de thématique.*

Une fois les points dérivants jugés utiles corrigés, vous pouvez supprimer tous ceux que vous avez ignorés.

10 Encoder un document texte

Sonal peut intégrer et encoder des documents au format texte, sans bande sonore ou vidéo associée. Cependant, les fonctionnalités d'encodage restent fortement influencées par la logique de découpage en "extraits" propre à Sonal. Le texte est vu comme sont vus les documents sonores dans Sonal : comme une continuité à scinder en plusieurs portions. A l'importation d'un document texte, Sonal le voit donc comme un seul grand extrait qu'il faudra découper.

- Spécificité du codage des documents texte

Le traitement des documents textuels dans Sonal repose sur une astuce : au lieu d'enregistrer les coordonnées en secondes des extraits, Sonal calcule la durée et la position relatives de chaque extrait *en pourcentage de la longueur totale (en mots)* et inscrit ces informations dans le fichier Rtr *comme s'il s'agissait de secondes*. Tous les fichiers textes ont donc une durée fictive de 999 secondes (99,9% de la durée) et ont des coordonnées calculées en proportion. Ainsi, un extrait qui occupe la première moitié d'un texte débutera à 0 « secondes » (0% en fait) et se terminera à 500 « secondes » (50%). Comme Sonal sait repérer les fichier texte, ces fausses secondes ne rentrent dans aucun calcul de durée et sont bien interprétées dans la fenêtre de codage dédié.

> *Cette spécificité induit cependant que, dans la fenêtre de corpus, lorsque les entretiens sont mis à l'échelle, tous les fichiers texte ont la même durée : 9999 secondes, soit un peu plus de 16 minutes. Une représentation corrigée sera proposée ultérieurement.*

Nouveauté importante de la V2

Dans la version 1, il était déconseillé de modifier fortement la longueur des extraits, d'allonger ou de raccourcir les textes. Le calcul des coordonnées était perturbé car les longueurs ne correspondaient plus aux pourcentages calculés.

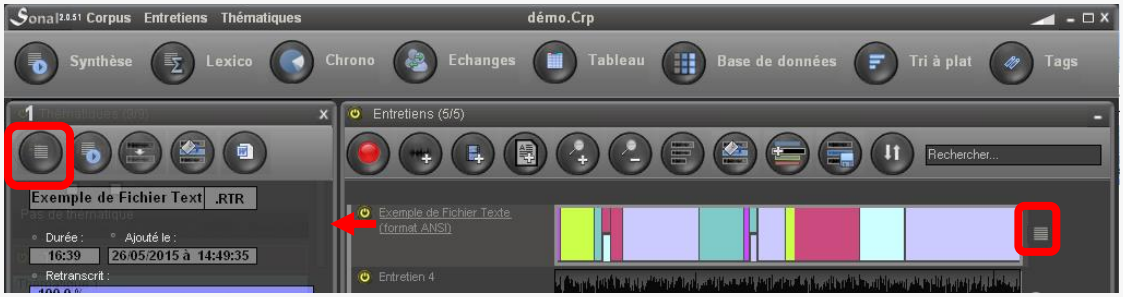
La version 2 corrige cette limite. **Vous pouvez modifier à volonté le contenu des extraits** dans la mesure où *toutes* les coordonnées des extraits sont recalculées à chaque modification.

10.1 La page de codage du texte

Le codage des fichiers texte se réalise dans une fenêtre dédiée. Son fonctionnement est très proche de celui de la page de codage des fichiers son, mais son maniement est globalement plus simple, car il n'y a pas de son à prendre en compte.

La principale différence avec la page de codage des enregistrements se situe dans la représentation globale de l'entretien, qui figure la nature textuelle du matériau, avec des pages juxtaposées latéralement. Le nombre de pages correspond au vrai nombre de pages du texte dans l'interface.

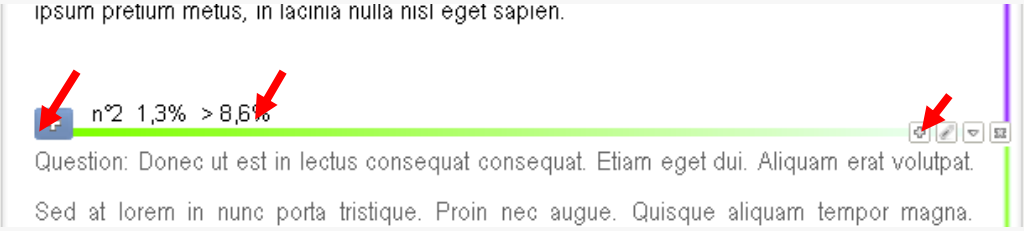
On y accède par les boutons ci-dessous :



- Utilisation

Le fonctionnement est relativement simple. En cliquant sur les pages situées en haut de l'écran, on progresse dans le document.

Pour éditer le contenu des extraits, on clique sur les coordonnées de l'extrait ou sur le gros "plus" à gauche de la bordure :

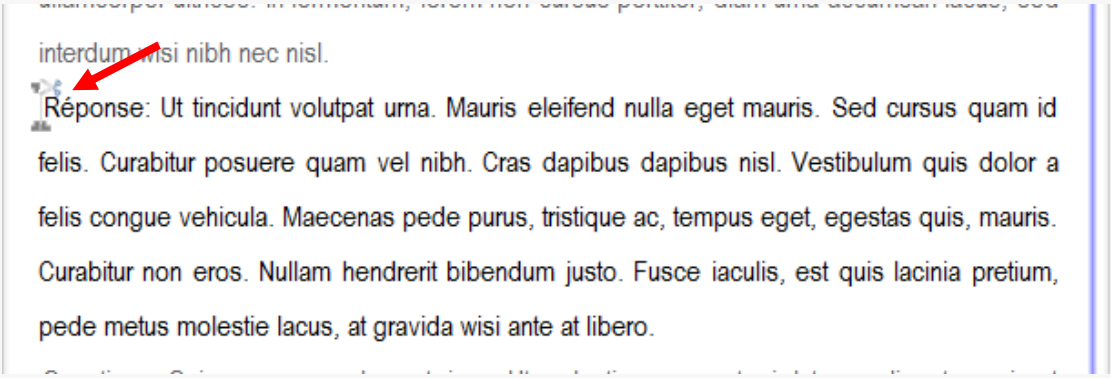


10.2 Découper des extraits dans les documents textuels

Pour encoder vos documents, il s'agira ensuite, comme dans la resynchronisation, de placer des bordures d'extrait dans le texte comme autant de fractures thématiques. Du premier extrait global on fait d'abord deux extraits, puis trois, et ainsi de suite.

Pour placer un point de partition, qui débute une portion thématique, on peut **cliquer sur le bouton "couper" en haut de page**, qui se mettra à clignoter, ou directement faire un **clic droit sur le mot voulu**.

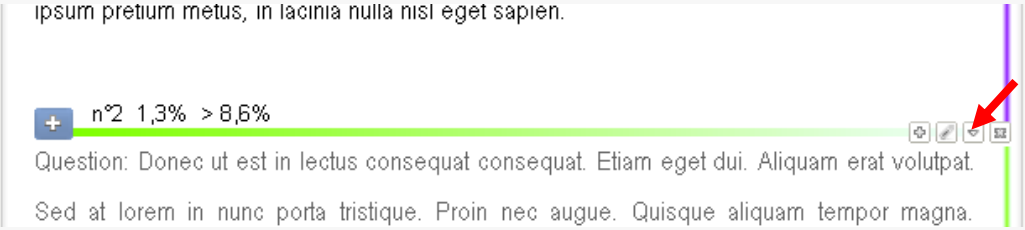
Au survol des mots du texte, le curseur prend alors la forme d'une paire de ciseaux, avec des bordures d'extraits fictives.



En cliquant sur le mot, vous scinderez l'extrait courant en deux parties.

10.3 Affecter une thématique à un extrait textuel

La manière la plus simple d'affecter une thématique à un extrait qu'on vient de couper est de cliquer sur le bouton ci-dessous :



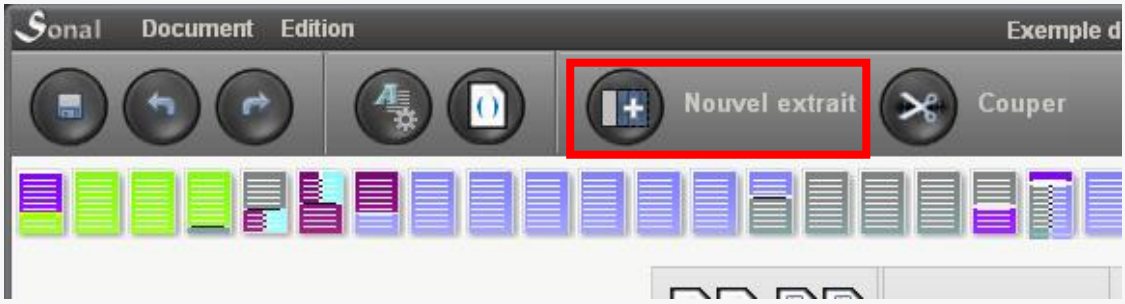
Une fois la thématique choisie, les pages correspondant à l'extrait sont coloriées.

Le fait de ne pas pouvoir superposer les extraits est contraignant mais on peut jongler avec cette logique d'encodage un peu particulière, qui permet finalement de tout faire assez simplement.

On peut par exemple superposer plusieurs thématiques par extrait pour prendre en charge les parties qui peuvent être encodées avec plusieurs thématiques :

10.4 Ajouter un nouvel extrait

Vous pouvez ajouter un nouvel extrait dans les documents textuels en cliquant sur le bouton ci-dessous :



Concrètement, cela signifie que Sonal va rajouter deux mots à la fin de votre texte (« Nouvel Extrait ») et les placer dans un nouvel extrait (dont les coordonnées risquent d'être entre 99,5 et 99,9%). Vous aurez donc un nouvel extrait tout en bas de votre page, que vous pourrez éditer et thématiser à volonté.

Plus il y a de mots dans un extrait, plus la réécriture du fichier est longue. Autrement dit, le temps de découpage est fonction de la longueur de l'*extrait* que vous découpez. Si vous découpez un texte très long, une solution pour gagner du temps peut être de pré-découper votre fichier en quelques grosses portions arbitraires, comme ça, l'extrait que vous découpez à chaque avancée du codage ne fera plus toute la longueur restante, mais seulement une partie et cela devrait aller beaucoup plus vite.

Concrètement, allez au premier quart du fichier (par exemple), coupez un nouvel extrait au hasard (inutile de mettre une thématique). Codez ce qui doit l'être en amont de la coupure que vous avez placée en avançant dans la lecture. Puis quand vous en arrivez au point que vous avez placé, annulez la partition que vous avez créée (cliquez sur la petite croix tout à droite).

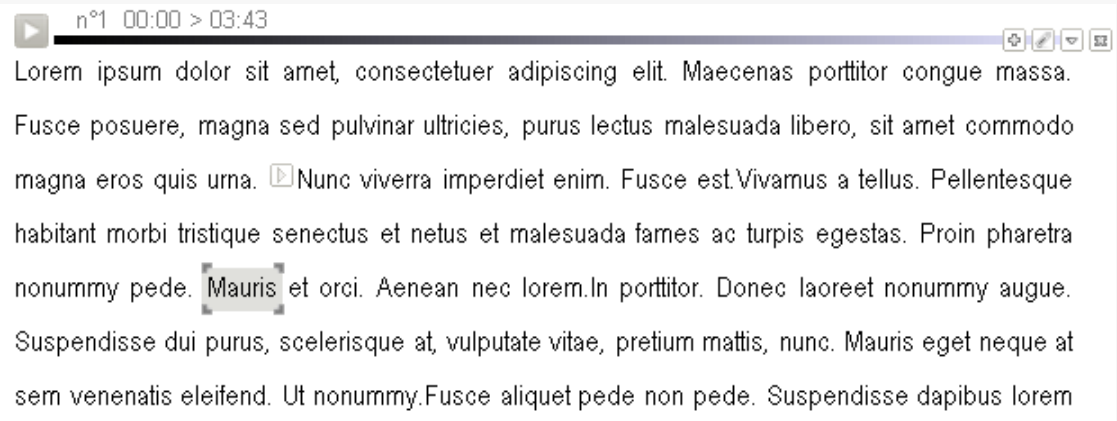
Refaites une nouvelle coupure plus loin. Et ainsi de suite...

11 La pondération qualitative

Une fois les extraits retranscrits, Sonal vous propose une dernière « couche » d'encodage, facultative, mais qui est susceptible de vous rendre service en phase de rédaction. Cette fonction se nomme « pondération qualitative ».

11.1 Marquer un passage

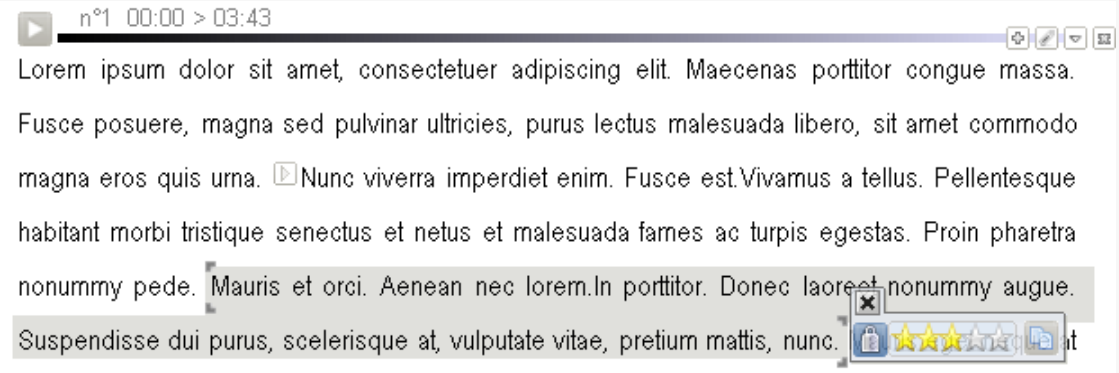
La pondération qualitative est disponible dans les fenêtres de codage (enregistrements et texte) et dans la fenêtre de synthèse. Son fonctionnement est simple. Il suffit de faire un clic gauche sur le premier mot de la portion à sélectionner. Celui-ci se grise :



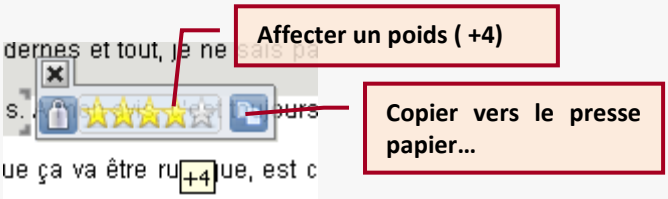
On survole ensuite le texte jusqu'au dernier mot de la portion à marquer, qui va se griser au survol.

Pour sortir de cette fonction, faites « échap » ou « esc » sur votre clavier.

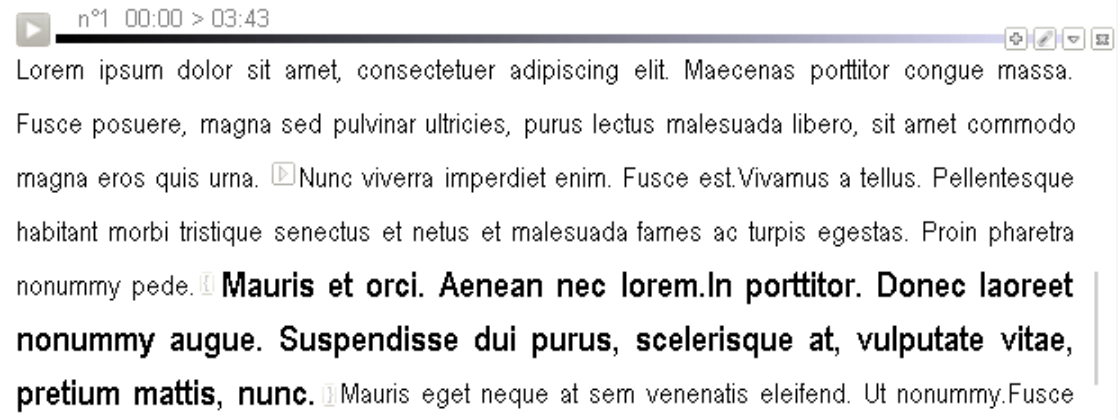
En cliquant sur le dernier mot de la portion voulue, un menu va apparaître :



Ce menu vous permet de faire deux choses : Affecter un poids à la partie sélectionnée ou la copier vers le presse papier. :



Une fois le poids conféré, le texte prend la forme voulue et la pondération est rappelée dans le corps du texte par des accolades discrètes et par un trait dans la marge:



11.2 Règles à connaître

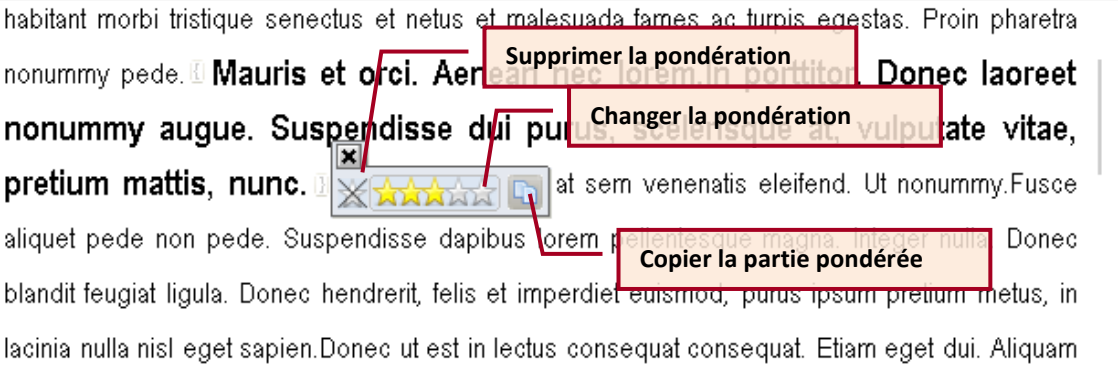
- On ne peut pas marquer une partie à cheval sur deux extraits

- Lorsqu'un mot est pondéré, on ne peut plus l'utiliser comme point de départ ou de fin pour une pondération. Cela empêche de débuter ou de terminer une sélection *à cheval* sur une autre. En revanche, une pondération peut en *englober* une autre. Par exemple, vous pouvez avoir toute une partie à +1 et seul un passage à l'intérieur de cette partie à +3 mais, pour obtenir ce résultat, il faut commencer par marquer le passage à +3 puis l'englober dans un marquage plus général à +1.

11.3 Modifier/annuler une pondération

- Changer le niveau de pondération ou la supprimer

Les marqueurs **{** et **}** placés dans le texte permettent d'afficher à nouveau le menu de pondération et de choisir un autre poids ou d'annuler tout simplement la pondération :



En cas de problème, rappelez-vous que vous avez toujours la possibilité d'aller directement dans le texte modifier ou supprimer les balises (voir ci-dessus).

Vous pouvez aussi, dans les pages de codage, demander le retrait de toutes les pondérations en allant dans le menu « balises » et en cliquant sur « effacer tous les formats ».

- Redéfinir la fin de la partie pondérée

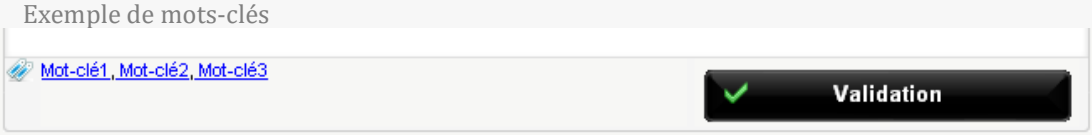
En faisant un **clic-droit sur le marqueur de pondération qui ferme la partie pondérée : « } »** , on recommence la sélection de la partie en conservant le point de départ.

12 Les mots-clés (tags)

12.1 A quoi servent les mots-clés ?

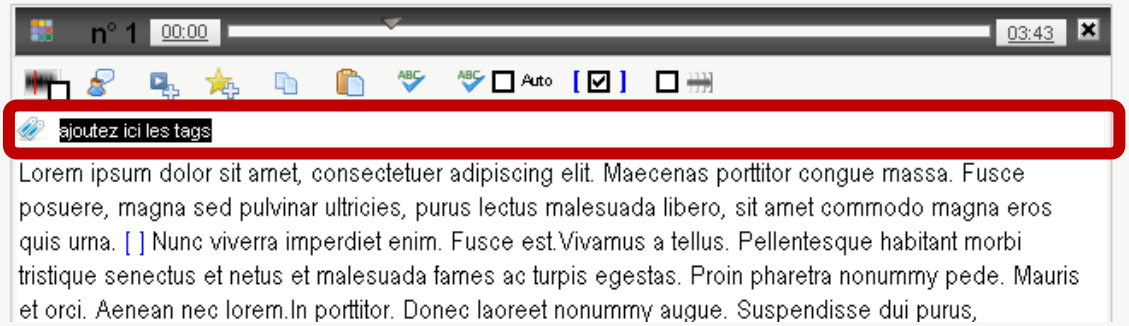
Chaque extrait créé dans Sonal peut se voir affecter une série (non limitée) de mots-clés destinés à le caractériser. L'objectif est de marquer tous les extraits qui partagent certaines caractéristiques qui ne sont pas renseignées ni dans la grille thématique, ni dans la base de données. Dans le cadre d'un entretien collectif, on peut par exemple mentionner qui parle etc. D'une manière générale, la forme des mots-clés est totalement libre et les espaces sont autorisés. Ils doivent seulement être séparés par une virgule :

Exemple de mots-clés

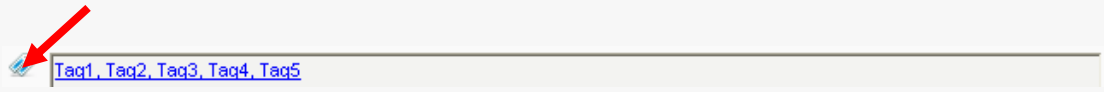


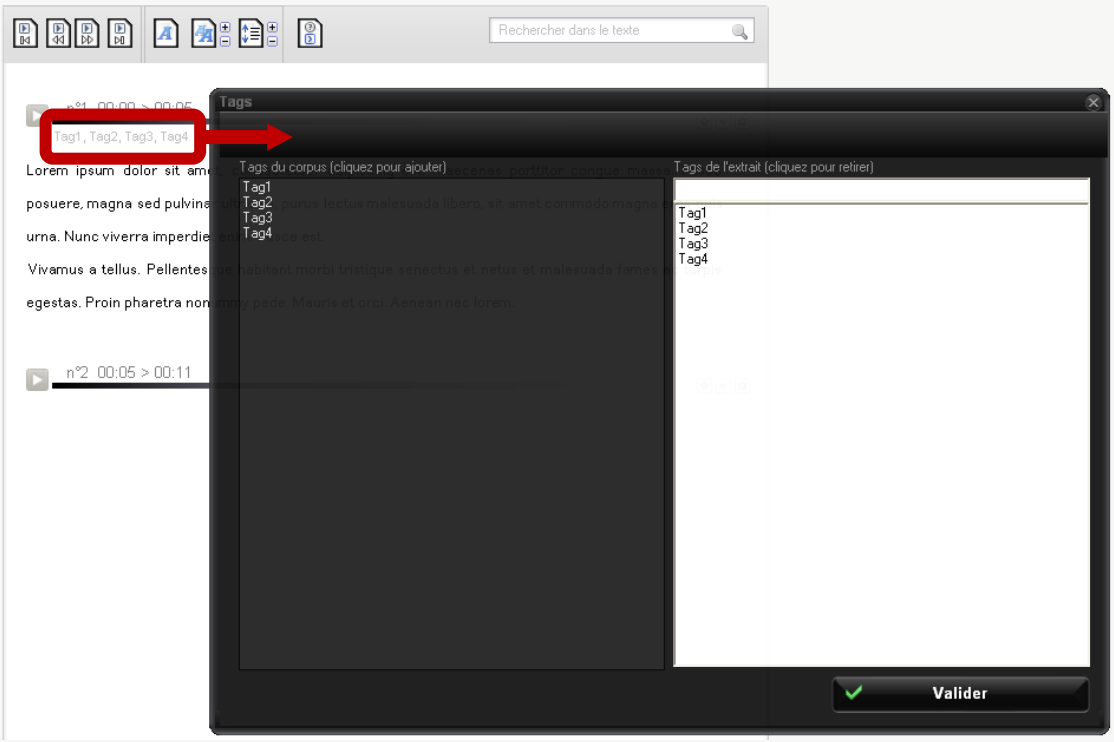
12.2 Ajouter des mots-clés

On peut ajouter des mots-clés depuis toutes les fenêtres de saisie du logiciel (page de codage, de synthèse). L'espace qui leur est dédié se trouve toujours juste au dessus de la zone de texte principale de l'extrait.



La saisie des tags est semi-automatique. Commencez à taper le mot-clé que vous souhaitez ajouter et celui-ci apparaîtra dans une liste sous la zone de saisie si vous l'avez déjà ajouté quelque part. En cliquant sur l'image qui symbolise les tags, vous pouvez utiliser la fenêtre d'inventaire des Tags, qui permet une gestion plus encadrée des mots-clés :



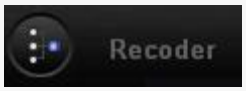


Le fonctionnement est simple, la liste à gauche contient l'intégralité des tags du corpus, la liste de droite ceux qui sont affectés à l'extrait. Cliquez sur les tags de la liste de gauche pour les ajouter, cliquez sur les tags de la liste de droite pour les retirer. Si vous souhaitez ajouter un tag qui n'existe pas encore, tapez-le dans la partie supérieure de la liste de droite et **validez avec entrée**.

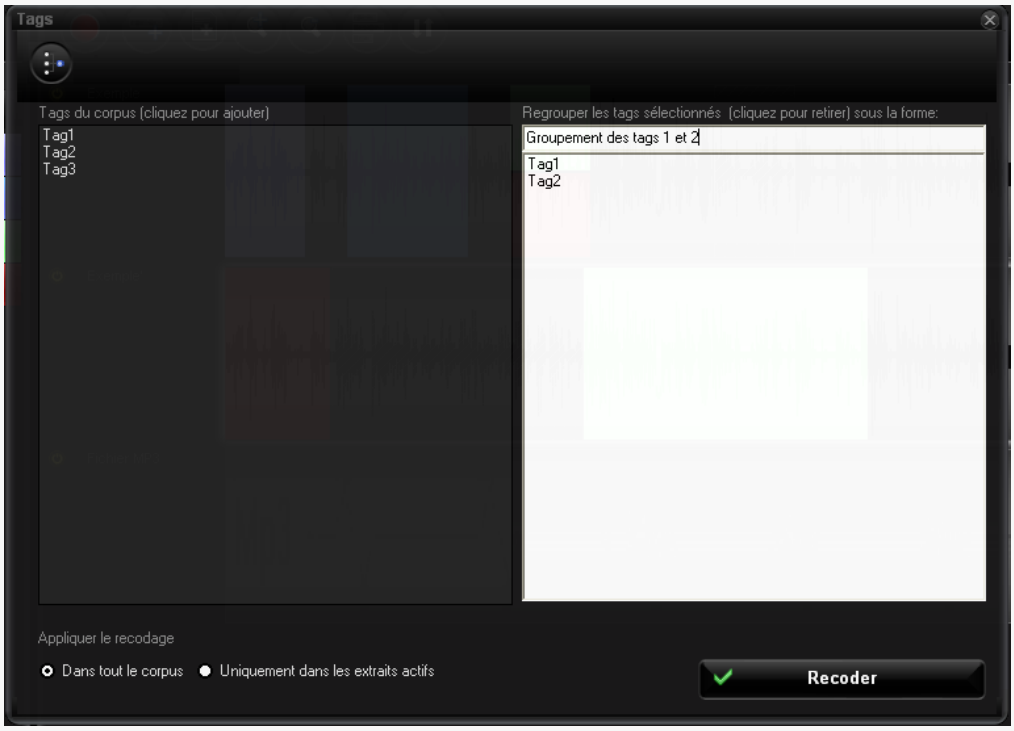
12.3 Recoder des mots clés

Il peut arriver que vous ayez besoin de regrouper plusieurs mots-clés sous un seul intitulé, ou que vous souhaitiez modifier un mot-clé et que celui-ci soit transformé dans tout le corpus.

Pour arriver à ce résultat, utilisez la fonction « recoder » accessible depuis la page des mots-clés :



L'interface qui s'ouvre alors à cette apparence, et fonctionne comme suit :



Les mots-clés sur lesquels vous voulez agir doivent être sélectionnés dans la liste de gauche. Ils apparaissent alors dans la liste de droite. Vous pouvez les en retirer en cliquant dessus. La zone de texte au sommet de la liste de droite vous permet de définir le nouveau nom à affecter au(x) mot(s)-clé(s) choisi(s).
Il ne vous reste qu'à définir si vous voulez que ce codage s'applique à tout le corpus ou aux seuls extraits actifs. Puis cliquez sur recoder. Après un message de confirmation, Sonal ira transformer les mots-clés des différents extraits retenus.

12.4 Filtrer les extraits par mots clés

Pour l'instant, le système de tri par mot-clé ne fonctionne qu'avec la synthèse. Une extension à toutes les fonctions d'analyse est prévue…

Pour filtrer les extraits par mots clés, cliquez sur le bouton "tags" dans la page de corpus. Dans la fenêtre qui apparaît, cliquez sur un mot clé pour voir apparaître tous les extraits qui le contiennent.
Attention, le filtrage par mot clé s'ajoute au filtrage général. Cela implique que si vous avez isolé une thématique, vous ne verrez que les extraits contenant le mot clé choisi et contenant la thématique donnée.

13 Exporter les retranscriptions

13.1 Les fichiers Rtr (bon à savoir)

Votre travail de retranscription est sauvegardé dans un fichier texte portant le nom de la bande avec pour extension Rtr (voir la partie consacrée au fonctionnement pour plus de détails).

Ce fichier s'ouvre avec le bloc note ou n'importe quel traitement de texte. *Vous pouvez donc récupérer le texte de vos retranscriptions en dehors de Sonal.*

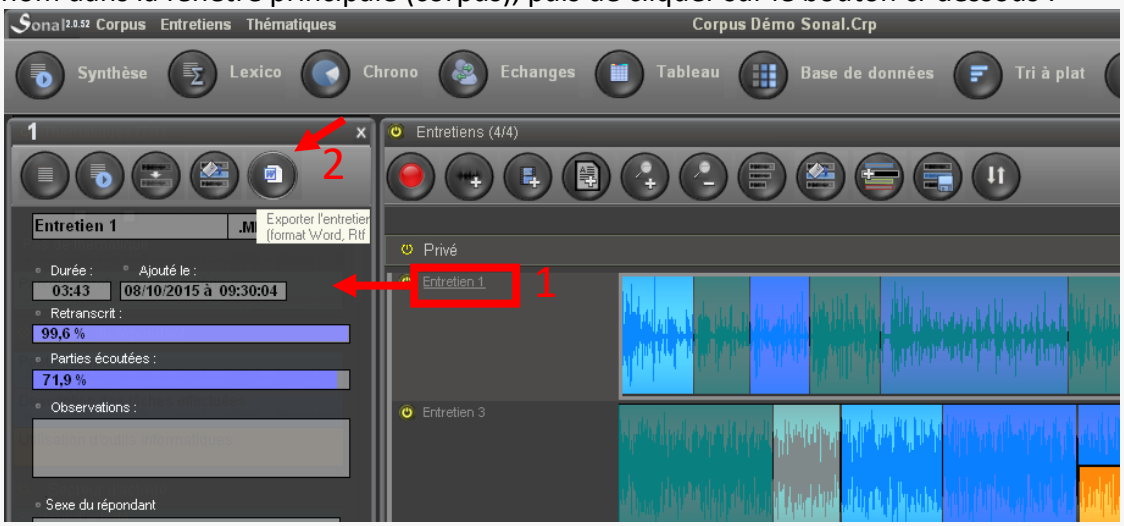
Le système de balisage de Sonal charge un peu le texte mais vous pourrez récupérer l'intégralité des retranscriptions.

Lorsque vous faites une copie de sauvegarde de votre corpus, Sonal enregistre tous vos fichiers Rtr dans un dossier à part.

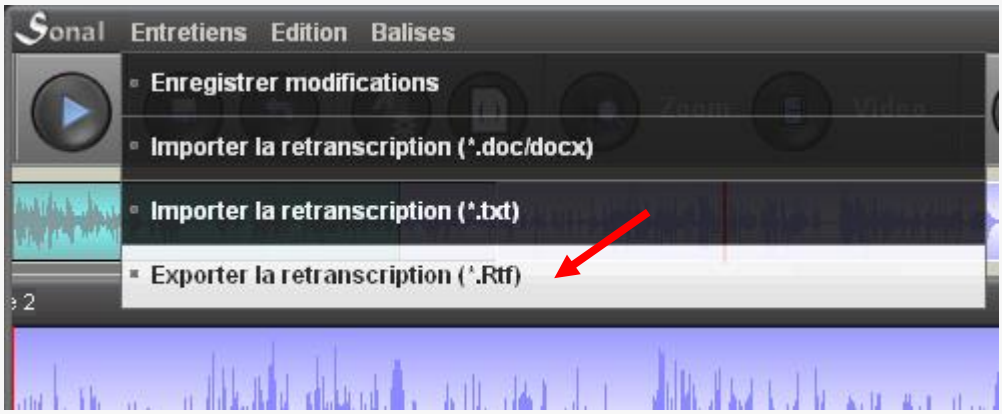
De même, à chaque enregistrement en cours de retranscription ou de codage, Sonal conserve dans le dossier « Restaur » les dix dernières versions du fichier Rtr modifié. Cela peut être utile pour récupérer des retranscriptions, même en cas de problème.

13.2 Exporter une retranscription

Pour exporter une retranscription d'entretien, le plus simple est de cliquer sur son nom dans la fenêtre principale (corpus), puis de cliquer sur le bouton ci-dessous :



Vous pouvez également afficher l'entretien dans la fenêtre de codage et de choisir le menu « Entretiens » puis « Exporter la retranscription » :



Dans les deux cas, un menu va apparaître, qui vous proposera différentes options pour l'export :

Les options disponibles sont les suivantes :

- Ouvrir dans Word

Sonal peut directement ouvrir Word et y envoyer la retranscription. A ce moment là, votre document n'a pas encore de nom, il faudra l'enregistrer dans Microsoft Word. C'est pratique dans un certain nombre de cas mais cela suppose que vous ayez Word

d'installé sur votre machine et que Sonal communique bien avec lui, ce qui n'est pas toujours le cas. L'export en Rtf est plus stable, car il n'est pas dépendant d'un logiciel externe.

- Exporter dans fichier Rtf

Pour un export en Rtf, il vous faudra d'abord définir le nom et l'emplacement du fichier texte que vous voulez exporter. Ceci fait, Sonal demandera que le fichier soit ouvert avec le traitement de texte par défaut du système d'exploitation. Cela vous permettra d'**utiliser open office** par exemple, ou n'importe quel logiciel compatible.
L'autre avantage du format Rtf est que vous pourrez **exporter les pondérations, ce qui n'est pas possible avec Word.**

- Exporter en html

L'export en Html 5 vous permet d'avoir un document texte lisible par votre navigateur Internet dans lequel vous pourrez également entendre le son associé, grâce à un lecteur intégré à la page, qui récupère les informations de synchronisation laissées dans le texte. C'est très pratique pour partager un document en ligne et permettre le même niveau de lecture que celui qu'offre Sonal. Là aussi, il faut commencer par définir le nom et l'emplacement du fichier Html que Sonal va créer.

Détails techniques : l'export html réagit différemment selon l'emplacement du fichier :
si le fichier est **dans un répertoire différent du corpus**, Sonal écrira dans le fichier html **l'adresse absolue des fichiers son** associés (ex : « C:\Mes Documents\Projets Sonal\Entretien1.Wav »). Cela fait que votre fichier html ne sera pleinement **utilisable que sur votre machine**. Ailleurs, l'adresse ne pointera sur rien.
Si vous enregistrez le fichier html dans le dossier du corpus, Sonal écrira l'**adresse relative** des fichiers (ex : « Entretien1.Wav »). En copiant le fichier Html et les enregistrements qu'il désigne dans un autre répertoire (voire sur un serveur en ligne), **le fichier Html fonctionnera partout**.

- Options d'affichage de l'export

Les différentes cases qu'il y a en dessous de la page permettent de choisir quoi exporter dans les retranscriptions.

- *Informations sur les interviews* : afficher le contenu de la base de données pour l'entretien en question
- *Observations sur les interviews* : afficher vos observations
- *Thématique des extraits* : voir les coordonnées temporelles et les thématiques des extraits
- *Tags des extraits* : voir les mots-clés affectés à chaque extrait
- *Afficher points de synchro*: voir les points de synchronisation (convertis en HH :MM:SS) dans le texte. Si vous exportez en Html, les points de

synchronisation seront convertis en boutons qui vous permettront de piloter la bande depuis le cœur du texte, comme dans Sonal.

- *Afficher formats* : représenter dans le texte les parties pondérées (*cette fonction n'est pour l'instant disponible que dans l'export en *Rtf.*)

Faites vos choix puis cliquez sur « exporter ».

13.3 Exporter le corpus vers Iramuteq

IRaMuTeQ est un logiciel d'analyse textuelle libre et gratuit qui propose des outils avancés qui permettent de prolonger ce que vous pouvez faire dans Sonal. Afin de faciliter le transfert de vos retranscriptions vers Iramuteq, une fonctionnalité a été ajoutée dans Sonal :

Illustration

Celle-ci assure le transfert de votre corpus dans un format lisible par Iramuteq, et notamment la mise en forme des métadonnées (variables de la base de données et thématique des extraits).
Sonal exporte tout le corpus dans un seul fichier texte, en respectant les normes d'iramuteq (voir mode d'emploi). Les variables ***publiques*** renseignées dans Sonal sont reportées en tête d'entretien et les thématiques au début de chaque extrait.

Attention, comme iramuteq ne supporte pas tous les caractères dans le texte, Sonal va expurger les noms de variable et de modalité de tous les espaces et les caractères non pris en charge. Pour faire simple, Sonal les supprime purement et simplement. Le résultat reste en général lisible mais cela peut vous jouer des tours (par exemple, « ministère » devient « ministre » car le « è » est supprimé.

Analyser un corpus

14 Organisation générale du système d'analyse

14.1 Choisir (ou pas) un angle d'approche des données

Sonal vous propose d'analyser votre corpus de plusieurs façons. Ces approches peuvent être utilisées en parallèle, tout au long de la construction du corpus. Chaque méthode partage avec les autres une logique commune : on va filtrer les extraits sur la base de leur thématique, du type d'entretien dont ils sont issus, de leur contenu textuel ou de leurs mots clés, avant d'analyser le résultat de ce filtrage de différentes manières.

- La première méthode, la plus intuitive est l'approche compréhensive qui consistera à envoyer les extraits triés dans une fenêtre de synthèse où ils pourront être relus et/ou réécoutés, et éventuellement exportés vers un traitement de texte. Le système de « pondération qualitative » permet en outre de repérer dans les différents extraits des passages jugés importants, qui pourront ensuite être affichés et exportés de manière isolée. C'est donc une logique d'inventaire et de synthèse des matériaux qui prime.

- La seconde approche, intrinsèquement quantitative, consiste à **étudier le vocabulaire utilisé dans les extraits filtrés**. On peut ainsi contraster le vocabulaire utilisé en fonction des grilles d'analyses renseignées dans la base de données et dans la liste des thématiques et savoir qui utilise plus spécifiquement quel vocabulaire sur quel sujet. **C'est l'analyse lexicométrique**.

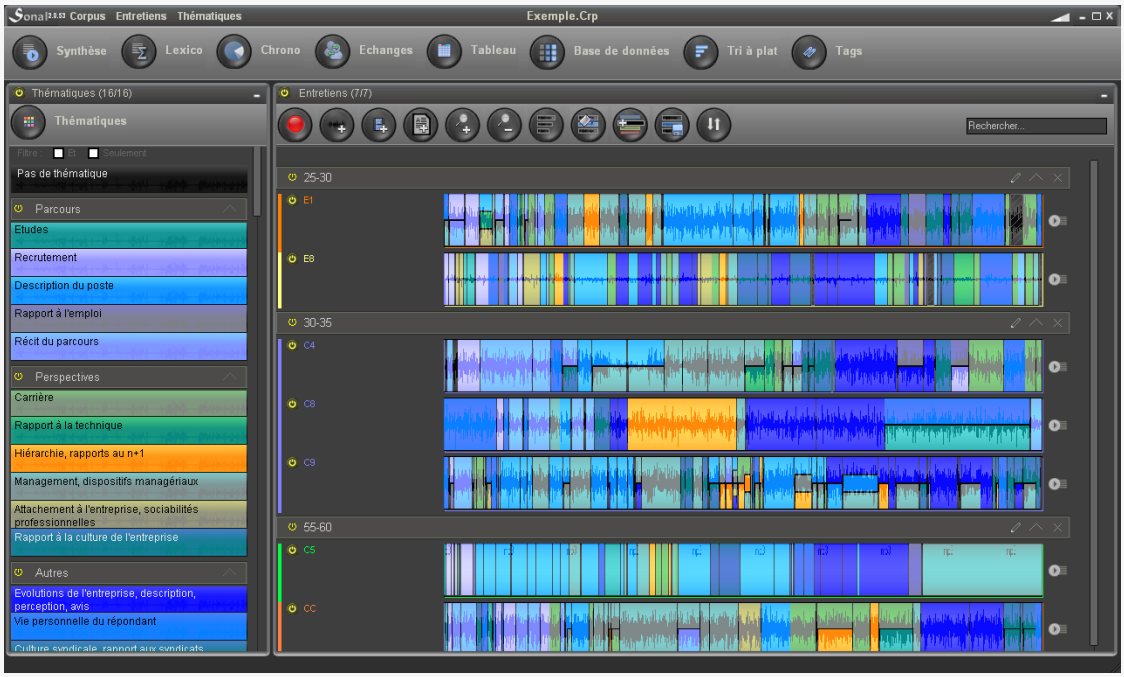
- La troisième approche consiste en **l'analyse des temps de parole**. De par le découpage de la bande lors du codage, il est possible de calculer le temps passé sur les différents thèmes par les différents (types de) locuteurs. **C'est l'analyse chronométrique.**

- La quatrième approche repose sur les balises de locuteurs et les points de synchronisation. Si les prises de paroles sont correctement renseignées, Sonal vous propose d'**analyser la dynamique des conversations** sur la base du nombre de mots utilisés par l'interrogateur et par ses répondants. C'est **l'analyse des interventions (ADI)**.

14.2 Le système de filtrage

Toutes les analyses proposées par Sonal reposent sur un même principe de filtrage. On va systématiquement isoler des sous-parties du corpus pour les lire, les étudier ou les exporter. Les outils d'analyse quantitative (sur les mots, les durées, ou les échanges) reposent toujours sur la comparaison du sous-ensemble défini par filtrage au reste du corpus. Cette méthode de filtrage repose sur plusieurs critères combinables, qui fonctionnent tous d'une manière similaire.

On peut voir un corpus encodé dans Sonal comme une sorte de « base de données qualitative » composée d'une collection d'extraits à la fois sonores et textuels. Ces extraits sont représentés par des "briques" de couleurs collées sur la bande et qui valent à la représentation synthétique du corpus le qualificatif de "mur" :



Chaque extrait possède des caractéristiques qui lui sont propres :

- des coordonnées temporelles spécifiques (les extraits ne se superposent pas)
- une chaîne de caractères (texte) correspondant à la retranscription

Et d'autres, qu'il partage avec d'autres extraits :

- une ou plusieurs thématique(s), qui définissent notamment sa couleur
- un entretien, duquel il est issu, et qui lui confère des caractéristiques renseignées dans la base de données.

Dans Sonal, un extrait c'est donc **1) une durée de bande, 2) du texte, 3) une thématique 4) une série de mots-clés et 5) des caractéristiques tirées de la base de données où sont enregistrées des informations sur l'entretien duquel est issu**

l'extrait. C'est sur la base de ces différents attributs que les extraits vont être activés ou non.

Caractéristiques des extraits						Caractéristiques de l'entretien			
Extrait n°	Début	Fin	Thématique	Texte de l'extrait	Tiré de l'entretien n° :	Sexe	Age	PCS	...
1	00:00:00	00:03:23	2	Question : Bonjour...	1	Femme	45-50	Employé(e)	...
2	00:03:23	00:05:57	3	Réponse : Oui effecti	1	Femme	45-51	Employé(e)	...
...	...	...	...	...	...	...	...	...	...

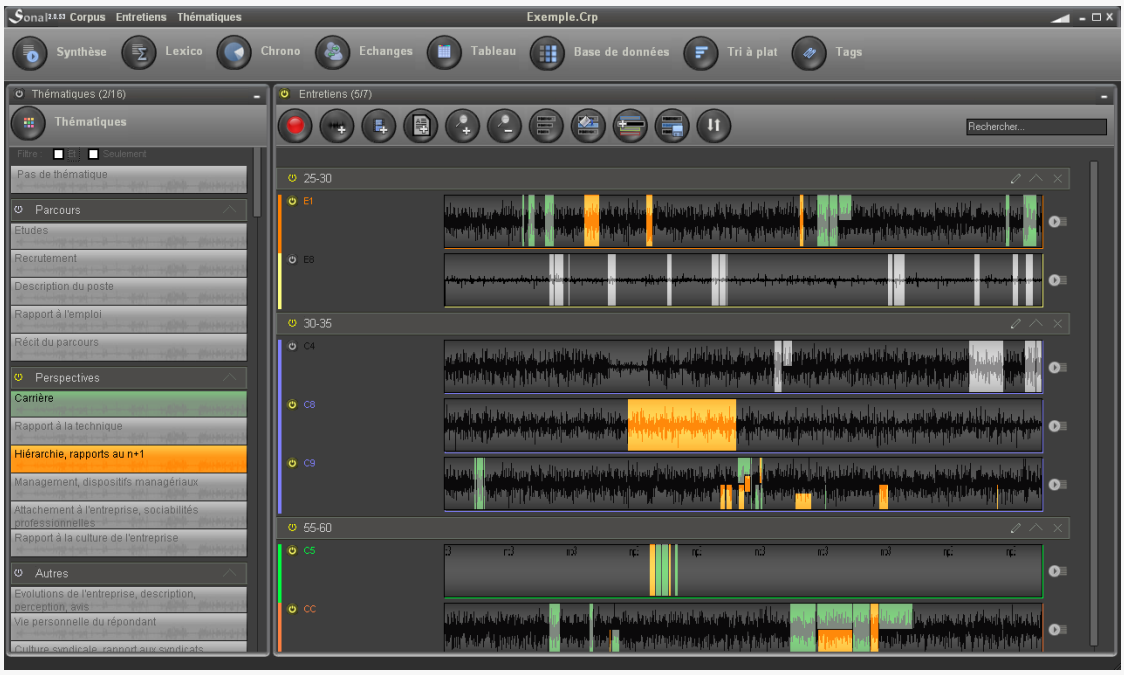
La première question est donc de savoir comment on active et comment on désactive les extraits.

Il y a **quatre approches possibles** et complémentaires pour activer ou désactiver les extraits :

- On peut choisir de ne garder actives **que certaines thématiques**
- On peut choisir de ne garder actifs que les **extraits contenant certains mots clés**.
- On peut choisir de ne garder actifs **que certains entretiens** en les choisissant individuellement
- On peut choisir de ne garder actifs **que certains types d'entretiens**, en se basant sur les caractéristiques de la base de données.

D'un point de vue graphique, Sonal offre une représentation très intuitive des filtrages puisque les analyses portent sur les extraits "allumés" repérables dans le "mur" du corpus.

Corpus filtré (certaines thématiques sont désactivées)



Il faut savoir que les filtrages opérés sont transversaux dans le logiciel. On peut filtrer les extraits depuis le corpus (cf. ci-dessus), mais on peut également demander des filtrages à l'intérieur des fenêtres d'analyse spécifiques (lexico ou chronométriques). **A chaque fois, les filtrages opérés à un endroit s'appliquent à tout le reste des interfaces.**

Hors exceptions, le système d'activation/désactivation des thématiques, des (types d') entretiens est partout le même :

Un clic gauche sur un élément conduit à l'activer ou à le désactiver.

Un clic droit sur un élément conduit à le maintenir seul activé

Le bouton situé au sommet des panneaux de filtrage permet l'activation ou la désactivation des toutes les catégories

- Filtrer par thématique

Pour allumer ou éteindre une thématique, un **clic gauche** sur la thématique considérée dans le volet thématique à gauche de la page suffira. Pour ne garder **que** cette thématique, utilisez un **clic droit**. Pour rallumer ou éteindre toutes les thématiques, cliquez sur le bouton on/off situé en haut à gauche du volet thématique. Dans la page de corpus, les extraits dont la thématique a été désactivée disparaissent.

Les « familles » thématiques créées par les séparateurs fonctionnent de même.

Il est également possible d'affiner la sélection pour gérer la thématisation multiple. Des cases à cocher "et" et "seulement" permettent de ne conserver lors du filtrage que les extraits qui contiennent soit plusieurs thématiques simultanément (fonction "et"), soit les extraits qui ne contiennent *que* la ou les thématiques activée(s), en excluant les extraits qui contiennent cette ou ces thématiques associées à d'autres.

- Filtrer par entretien

Les entretiens s'allument et s'éteignent individuellement en cliquant sur le bouton on/off situé à l'extrême gauche de la ligne de l'entretien dans le corpus. Là aussi, le clic droit revient à isoler l'entretien en question et le bouton on/off situé en haut du volet consacré aux entretiens permet l'activation/désactivation collective des entretiens. Les extraits d'un entretien désactivé restent visibles mais sont grisés.

- Filtrer par variables

Le filtrage par variables revient au même que le filtrage par entretien mais procède de manière collective. Au lieu d'activer ou de désactiver les entretiens individuellement, vous pouvez le faire au regard de leurs caractéristiques. Cette fonction passe par l'affichage du tri à plat, qui va vous donner la composition de votre corpus pour les différentes variables.

Après avoir déplié les variables, le clic gauche vous permettra d'allumer ou d'éteindre les entretiens. *Attention, ici, le clic droit vous permettra de modifier l'intitulé de la modalité, pas de l'isoler.* Le bouton on/off situé en haut du volet vous permettra de remettre le filtrage à zéro.

Filtrage des entretiens en fonction de leurs attributs

En cumulant ou non les caractéristiques de filtrage, vous pourrez donc très facilement sélectionner une collection d'extraits que vous pourrez soumettre à l'une des approches proposées par le logiciel

- Filtrer les extraits par leurs mots-clés.

Vous avez également la possibilité de filtrer les extraits par mots-clés. Le fonctionnement est un peu différent : il faut afficher la fenêtre des mots clés (bouton « Tags » - voir ci-dessous) :

Vous verrez alors apparaître une liste de mots-clés : plus un mot clé est fréquent, plus il est gros. En cliquant sur un mot clé, on accède directement à la fenêtre de synthèse où sont compilés tous les extraits qui contiennent le mot clé en question.

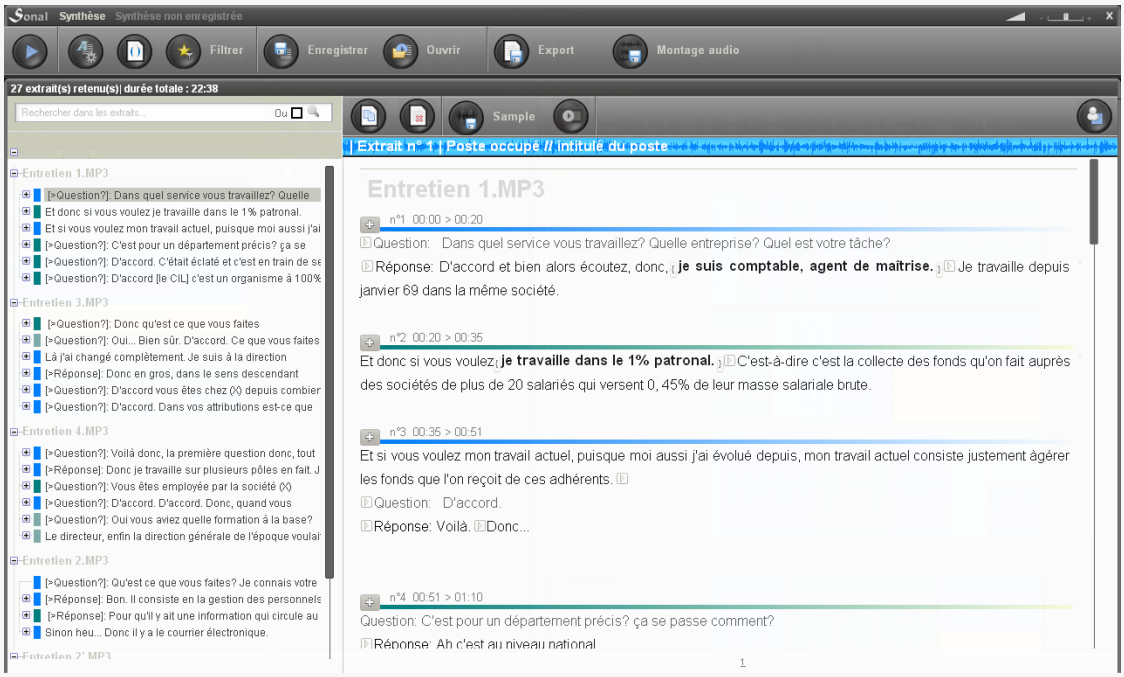
NB : Cochez ou décochez la case « filtrer les extraits » pour que les autres critères de filtrage restent opérants ou non.

15 L'analyse qualitative (synthèse)

Une fois un filtrage des extraits opéré, Sonal propose d'abord une approche très simple des matériaux, consistant à faire l'inventaire de tous les extraits dont on dispose sur un sujet donné.

Pour obtenir cette liste, cliquez simplement sur le bouton "synthèse" en haut à gauche de la page du corpus :

La fenêtre qui apparaît alors vous permettra de relire et/ou de réécouter tous les extraits sélectionnés.



La colonne de gauche contient tous les extraits retenus. Lorsque l'on clique sur l'un d'entre eux, la bande son se lance et le texte de l'extrait apparaît dans la partie droite de la page. Une fois un extrait terminé, Sonal lance automatiquement le suivant.

> *Si l'extrait contient des points de synchronisation, un petit bouton + apparaît sur la gauche de l'étiquette d'extrait. Au clic, la liste des points apparaît sous l'étiquette de l'extrait. Vous pouvez ainsi accéder plus rapidement à certains passages.*

Libre à vous d'utiliser par la suite le cadre interprétatif qu'il vous plaira; Sonal se contente de vous mettre à disposition l'inventaire exhaustif des extraits portant sur un sujet donné et/ou étant issus de certains (types) d'entretiens. On parle donc d'analyse qualitative car elle repose sur une démarche interprétative mais il n'y a pas à proprement parler de méthode. Il s'agit tout simplement de s'assurer qu'on ne rate aucun passage intéressant.

Une fois la synthèse réalisée, il est possible de copier des extraits individuellement pour les coller dans un traitement de texte, de supprimer de la synthèse ceux qui ne présentent pas d'intérêt etc.

15.1 Modifier les extraits depuis la synthèse

La fenêtre de synthèse sert avant tout à compiler et exporter des extraits filtrés, mais c'est aussi une fenêtre permettant l'édition des extraits. Cette fonctionnalité peut s'avérer extrêmement utile dans certaines situations.
Lorsque vous ouvrez un extrait, la touche entrée de votre clavier (ou le bouton + ci-dessous) vous permet de l'éditer pour modification :

La fenêtre de synthèse peut donc également être utilisée pour faire des modifications sélectives dans votre corpus.

Utile, vous pouvez également rouvrir l'entretien duquel est issu votre extrait pour le voir « en contexte ». Pour ce faire, cliquez sur le bouton ci-dessous :

C'est un autre moyen de modifier les extraits.
Vous pouvez également appliquer et modifier les pondérations depuis la synthèse.

Attention la prise en compte des modifications dans les pondérations n'opère immédiatement que pour l'extrait sélectionné. Bien que l'on puisse également agir sur les extraits qui suivent, les pondérations risquent de ne pas apparaître immédiatement.

15.2 Rechercher un mot clé

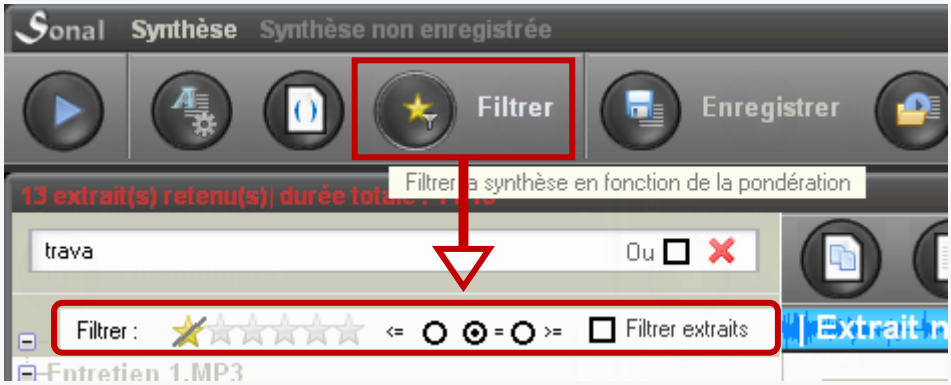
Cette fenêtre de synthèse vous permettra également de faire des recherches lexicales à l'intérieur des extraits. Tapez dans la zone de recherche le mot que vous recherchez et les extraits seront à nouveau filtrés en fonction de leur contenance.

La ligne de chaque extrait contient le premier contexte d'apparition du mot puis, si vous cliquez dessus, apparaît une liste des contextes dans lesquels le mot se trouve. Si vous cliquez sur ces lignes, vous êtes amenés à leur position dans l'extrait, ce qui est pratique si vous réalisez des extraits très longs.

> *Cette fonctionnalité vous permet également de rechercher un mot clé à l'intérieur de tout le corpus. Il suffit de ne pas filtrer les extraits avant synthèse.*

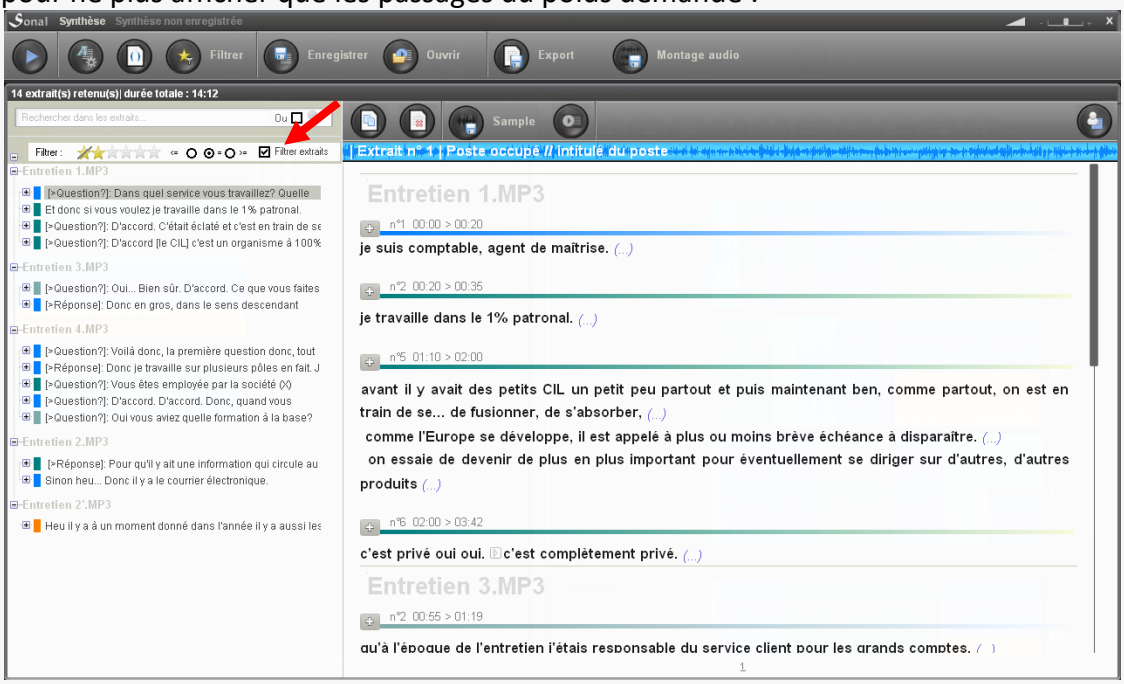
15.3 Filtrer les extraits par le « poids » de leur contenu

Le système de pondération qualitative peut être utilisé pour filtrer les extraits retenus en fonction de leur contenu. Pour afficher ou masquer le panneau de filtrage, utilisez le bouton en forme d'étoile :



Puis sélectionnez un niveau de filtrage déterminé par le nombre d'étoiles. Lorsqu'un niveau est défini, n'apparaissent plus que les extraits qui contiennent au moins un passage pondéré au niveau donné.

Si vous cliquez sur la case « filtrer extraits », le contenu des extraits retenus est filtré pour ne plus afficher que les passages du poids demandé :

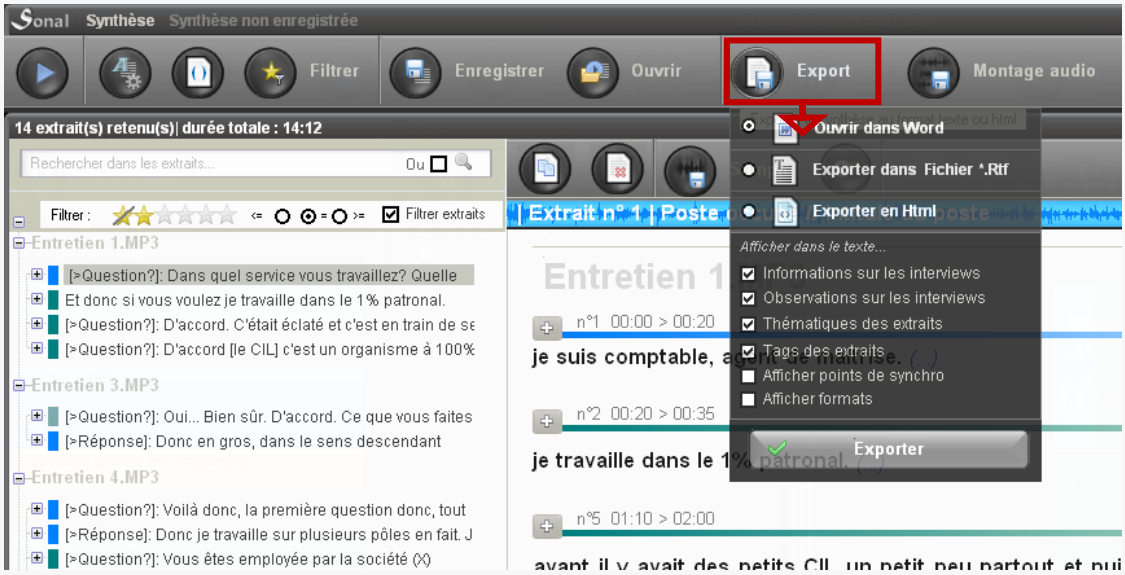


Cela peut être utile pour exporter un recueil de citations jugées illustratives vers un traitement de texte en cours de rédaction.

15.4 Exporter le résultat d'une synthèse

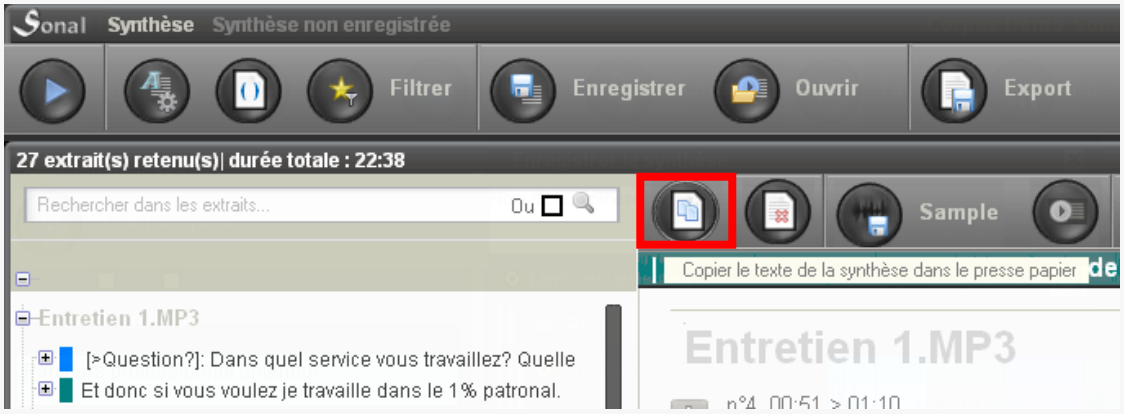
- Exporter le texte

Lorsque vous avez réuni une collection d'extraits qui vous semblent dignes d'intérêt pour illustrer tel ou tel phénomène relatif à votre objet d'étude, vous pouvez exporter le contenu de cette synthèse dans Word ou dans n'importe quel traitement de texte en cliquant sur le bouton d'export :



Les options sont les mêmes que pour l'export des entretiens entiers. Voir ici.

Si vous voulez n'exporter que le contenu textuel d'un extrait seul, cliquez sur le bouton ci-dessous :



- Exporter le son

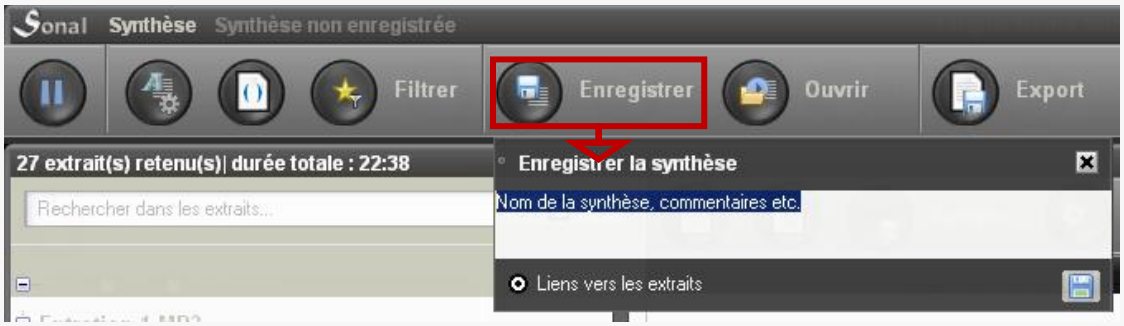
Vous pouvez également choisir d'exporter le son. Si vous cliquez sur le bouton « montage » ci-dessous, Sonal réalisera un montage des différents extraits sonores retenus pour faire un seul fichier son.

Attention, ce montage n'est possible que si tous les fichiers son sont de même format et que ce format est soit WAV, soit MP3.

Vous pouvez également choisir de « sampler », ou échantillonner les extraits individuellement, c'est-à-dire de découper chaque extrait et d'en tirer un seul et unique fichier son. Pour ce faire, cliquez sur le bouton ci-dessous :

15.5 Enregistrer une synthèse

Vous avez également la possibilité de sauvegarder votre synthèse à l'intérieur de Sonal, pour la rouvrir ultérieurement. Pour sauvegarder une synthèse, cliquez sur le bouton ci-dessous et ajoutez des commentaires qui vous permettront de la caractériser :



Après enregistrement, votre synthèse sera sauvegardée dans un fichier portant l'extension *.snt, situé dans un sous-dossier du dossier qui accueille votre corpus nommé « Synth ». Ce fichier ne contient aucune donnée en elle-même, seulement des informations permettant d'extraire à nouveau les données du corpus. Si vous modifiez un extrait, celui-ci apparaîtra modifié après réouverture de la synthèse.

15.6 Ouvrir une synthèse enregistrée

Le bouton ci-dessous vous permet de rouvrir toutes les synthèses enregistrées pour votre corpus.

16 L'analyse lexicométrique

Pour lancer une analyse lexicométrique, **cliquez sur le bouton « Lexico » situé dans la page de corpus.**

16.1 A quoi sert l'analyse lexicométrique?

- Compter les occurrences des différents mots (ou "formes")

L'analyse lexicométrique permet d'abord un **décompte des mots les plus fréquemment utilisés dans votre corpus** afin de repérer, au-delà des formes généralement très banales qui occupent le haut de la liste, les termes particuliers les plus récurrents dans vos entretiens. La spécificité de Sonal est que vous pourrez non seulement relire mais aussi **réécouter tous les passages qui contiennent les mots qui vous intéressent.** Dès lors, la lexicométrie joue un rôle d'indexation de la bande sonore et permet une entrée différente dans les matériaux, qui transcende les catégories thématiques et les caractéristiques des entretiens.

- Etudier les spécificités langagières associées à un sujet et/ou à un (type de) locuteur etc.

L'analyse lexicométrique prend tout son intérêt lorsqu'elle ne porte plus sur l'intégralité du corpus, mais sur une sous-partie, que vous aurez préalablement filtrée. En pareil cas, Sonal vous permettra de repérer très rapidement le vocabulaire le plus caractéristique de cette sous-partie. Par exemple, il devient possible d'étudier les mots les plus utilisés par un type de répondant sur un sujet donné, en contrastant en fonction des caractéristiques possibles. Vous accédez ainsi aux différences de

formulation sur un sujet donné par différents groupes de locuteurs, ce qui cadre avec une tradition structurale de l'analyse lexicale dans laquelle le vocabulaire est le reflet des positions sociales.

- Etudier les catégories de répondants ou les sujets dans lesquels une forme est la plus fréquente.

Sonal propose une approche complémentaire en vous permettant, pour un mot donné, de rechercher dans quelles partitions du corpus il est le plus présent et le plus spécifique. On peut donc faire des allers-retours entre une entrée par les catégories (quel vocabulaire est le plus lié à tel sujet par exemple) et une entrée par le vocabulaire (dans quelle sous population du corpus trouve-t-on le plus fréquemment et le plus spécifiquement un mot donné?)

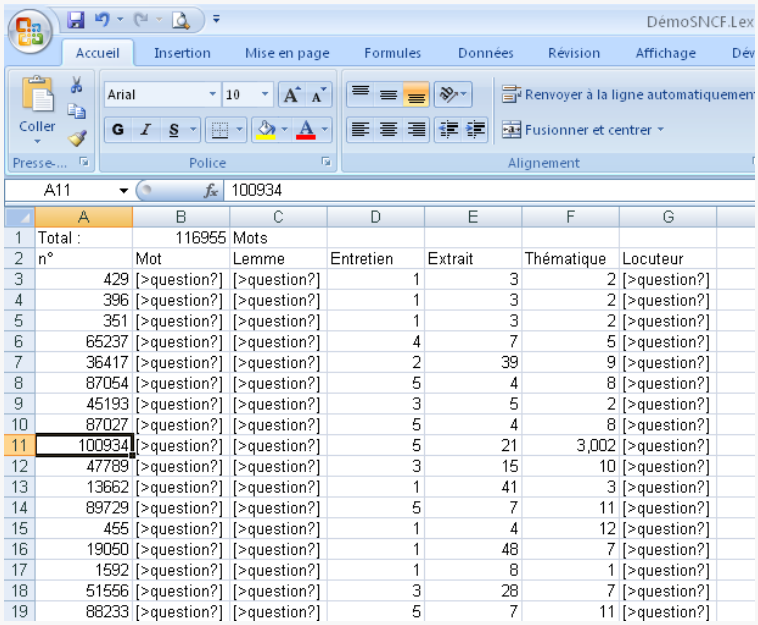
16.2 La base lexicale

Les analyses lexicométriques sont réalisées à partir d'une base lexicale lisible par excel ou calc et qui contient les différents mots du corpus, avec les informations clés sur les extraits dont ils sont issus (rang, mot, lemme, entretien, extrait, thématique, locuteur).

Ce fichier est placé dans le répertoire qui contient le corpus et porte le nom du projet suivi de « Lexico.xls ».

> *Malgré son extension, ce fichier est seulement* lisible *par Excel, c'est un simple fichier tabulé qui comporte une extension «xls » pour permettre son ouverture automatique avec le tableur installé sur la machine.*

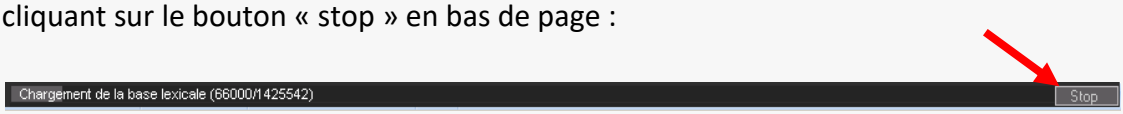
Base lexicale



	A	B	C	D	E	F	G
1	Total :	116955	Mots				
2	n°	Mot	Lemme	Entretien	Extrait	Thématique	Locuteur
3	429	[>question?]	[>question?]	1	3	2	[>question?]
4	396	[>question?]	[>question?]	1	3	2	[>question?]
5	351	[>question?]	[>question?]	1	3	2	[>question?]
6	65237	[>question?]	[>question?]	4	7	5	[>question?]
7	36417	[>question?]	[>question?]	2	39	9	[>question?]
8	87054	[>question?]	[>question?]	5	4	8	[>question?]
9	45193	[>question?]	[>question?]	3	5	2	[>question?]
10	87027	[>question?]	[>question?]	5	4	8	[>question?]
11	100934	[>question?]	[>question?]	5	21	3,002	[>question?]
12	47789	[>question?]	[>question?]	3	15	10	[>question?]
13	13662	[>question?]	[>question?]	1	41	3	[>question?]
14	89729	[>question?]	[>question?]	5	7	11	[>question?]
15	455	[>question?]	[>question?]	1	4	12	[>question?]
16	19050	[>question?]	[>question?]	1	48	7	[>question?]
17	1592	[>question?]	[>question?]	1	8	1	[>question?]
18	51556	[>question?]	[>question?]	3	28	7	[>question?]
19	88233	[>question?]	[>question?]	5	7	11	[>question?]

Lorsque vous cliquez sur le bouton Lexico dans la page du corpus, si vous avez déjà créé cette base, le logiciel la chargera, sinon il en écrira une nouvelle, ce qui prendra plus de temps. **La durée de chargement est variable en fonction des dimensions du corpus et de la puissance de votre ordinateur**. Pour un très gros corpus, le chargement peut prendre une à plusieurs minutes.

Si vous avez lancé une lexico par erreur, vous pouvez interrompre le processus en cliquant sur le bouton « stop » en bas de page :



Si vous avez le moindre problème avec la base lexicale et qu'elle vous empêche d'accéder à la fonction de lexicométrie, il vous suffit de supprimer le fichier « lexico.xls » dans le dossier projet. Au prochain rechargement, une base sera régénérée.

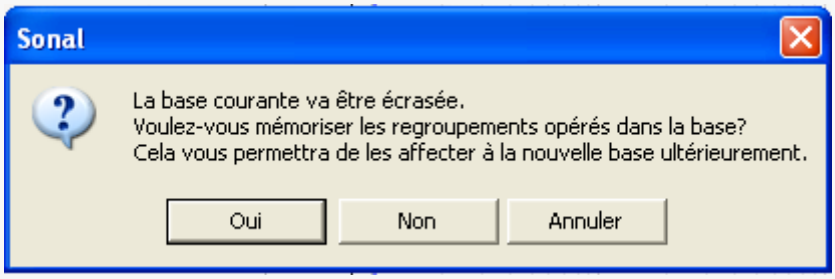
Si Sonal rencontre un problème au chargement de la base (ce qui est fréquent si vous modifiez la composition de votre corpus, par exemple en ajoutant ou retirant un entretien), il vous proposera d'en écrire une nouvelle.

- Réinitialiser la base lexicale

Vous pouvez également choisir de vous-même de réinitialiser votre base avec le bouton :

situé en bas à gauche de la page de lexicométrie.

Sonal réécrira alors intégralement la base et toutes les lemmatisations seront effacées. Si vous avez déjà travaillé sur votre base, il est préférable de répondre favorablement à la question qui vous est posée :

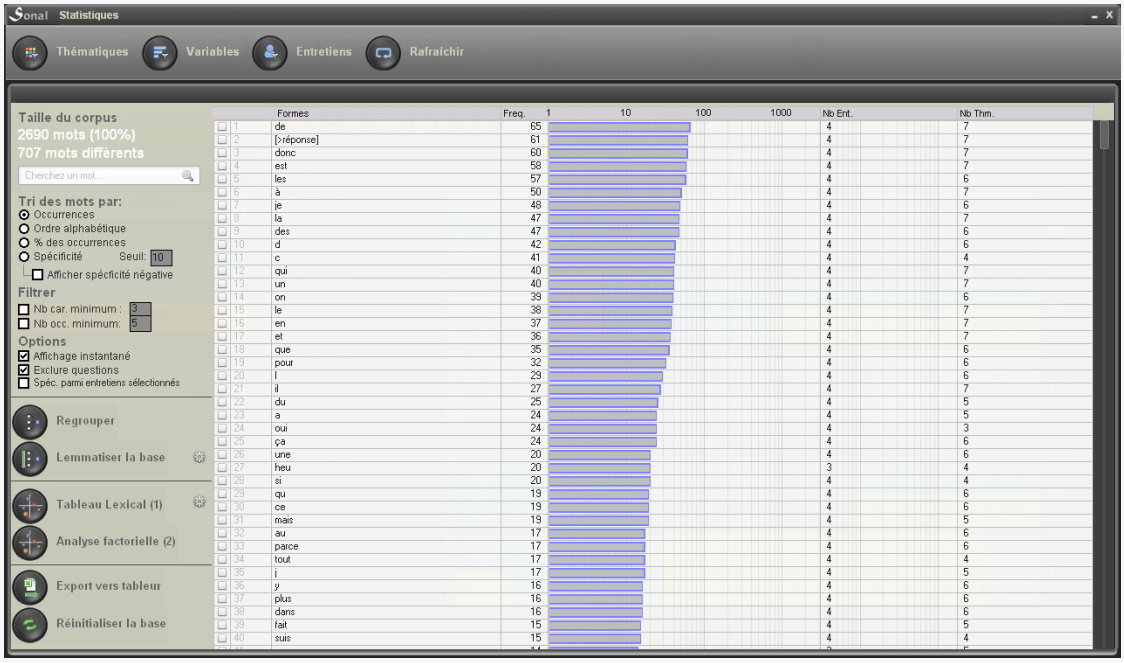


Vous aurez alors la possibilité de recharger les lemmes crées dans la nouvelle base.

16.3 La page de lexicométrie

Vous arrivez donc sur une page qui renseigne sur les occurrences des différents mots présents dans les extraits retenus.

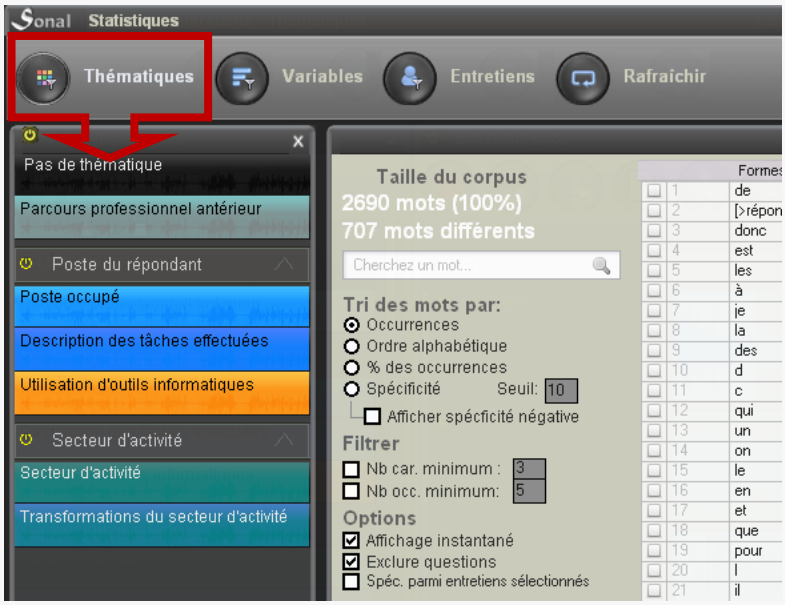
Fenêtre de lexicométrie



Avant de se pencher sur la signification des graphiques et des indicateurs affichés, il faut d'abord faire le tour de l'interface…

- Les panneaux de filtrage

Pour afficher un mode de filtrage particulier (par thématique, par variable ou par entretien), cliquez sur l'un des trois boutons dans la partie haute de la page. Dans l'exemple ci-dessous, un filtrage par thématiques a été réalisé.

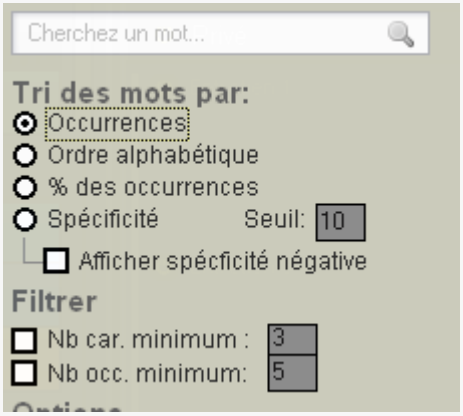


16.4 Le tri des mots

Dans la partie droite de la page, apparaissent des listes de mots qui sont déterminées par le mode de tri utilisé, et par les filtres introduits.

- Le tri par occurrences

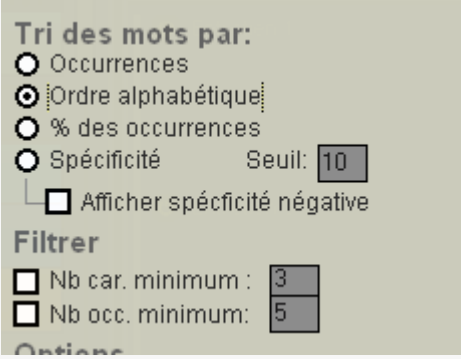
Le tri par occurrences affiche la **fréquence des différents mots par ordre décroissant** à l'intérieur de la partie du corpus sélectionnée.



Des barres bleues représentent ces fréquences suivant une échelle logarithmique. C'est l'occasion de vérifier la fameuse loi de Zipf, que traduit (normalement) la décroissance très progressive de la fréquence d'apparition des formes à l'intérieur d'un corpus.

- Le tri par ordre alphabétique

Le tri par ordre alphabétique affiche les mêmes informations que le tri par occurrences. Il ne s'en différencie évidemment que par le tri effectué. A noter : les nombres précèdent les lettres.

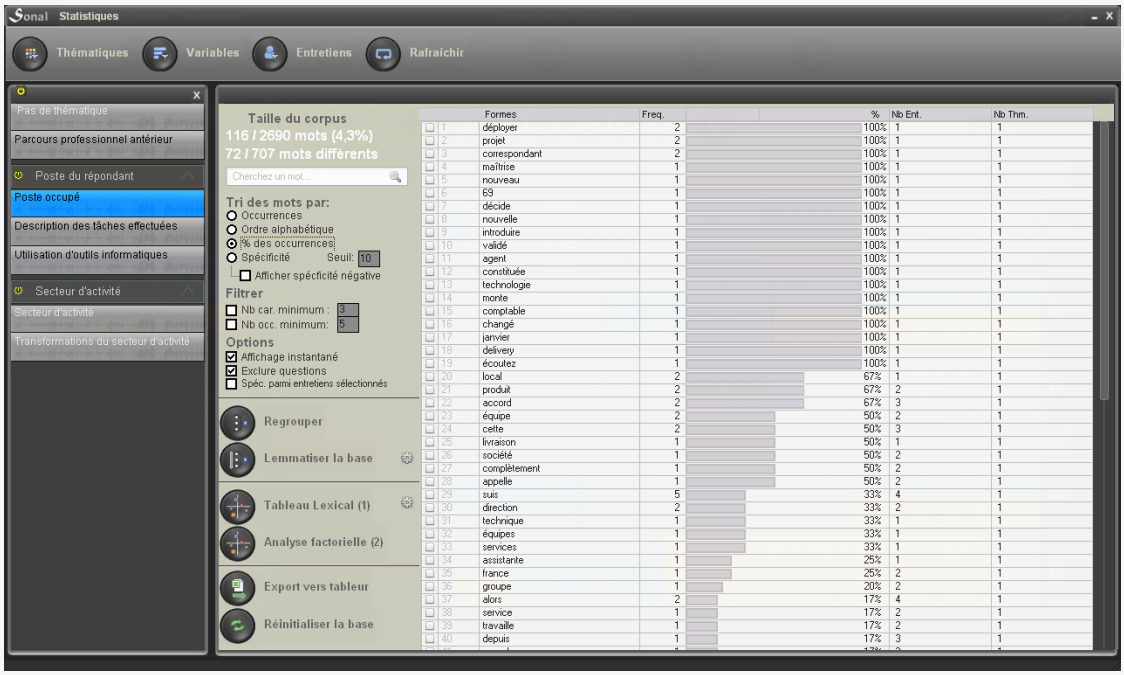


- Le tri par pourcentage des occurrences

Le tri par occurrences vous permettra de **trouver rapidement les formes qui sont spécifiques aux extraits que vous avez isolés**. Les mots 100% spécifiques, qui occupent le haut de la liste, n'apparaissent que dans les extraits sélectionnés, et nulle part ailleurs dans le corpus. Les mots à 50% sont pour moitié dans les extraits, pour moitié dans le reste du corpus.

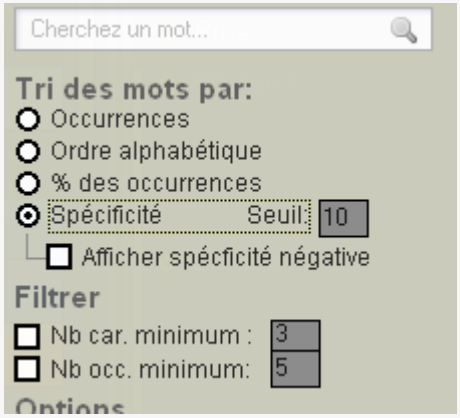
Ce tri n'a de sens que si vous avez sélectionné une sous-partie du corpus. **Si tous les extraits sont actifs, ce tri est sans effet**.

La liste représente les formes triées d'abord par pourcentage de spécificité, puis par occurrences, ce qui veut dire que vous avez au sommet de la liste les formes exclusives les plus fréquentes :



- Le tri par spécificité

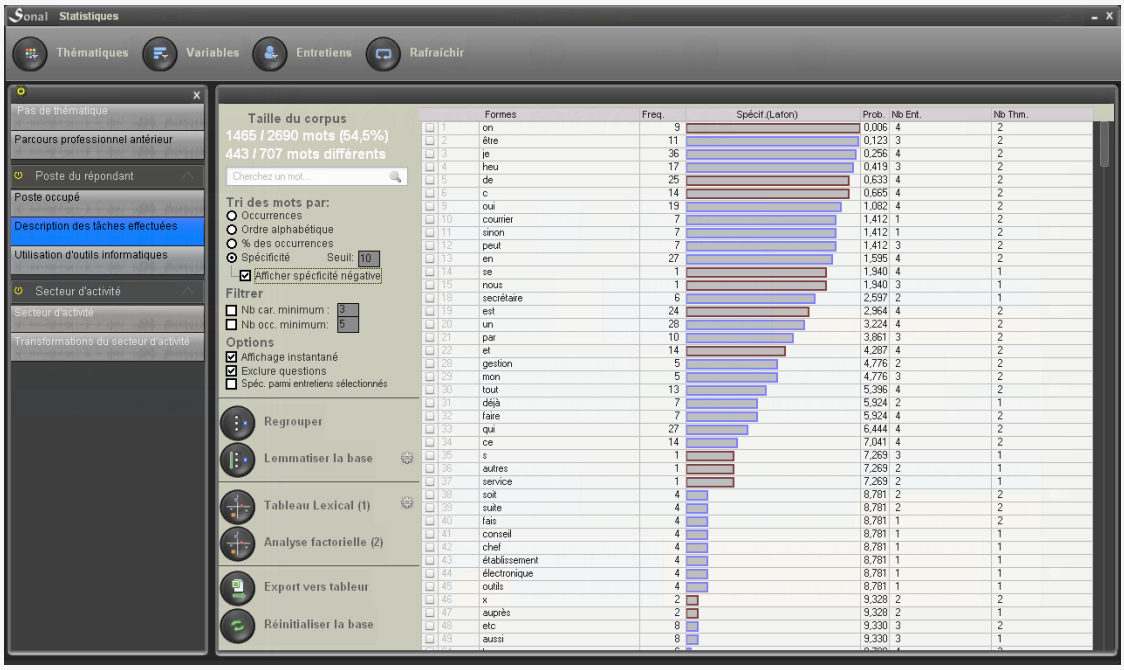
Le tri par spécificité est un outil dédié à l'analyse des particularités langagières des extraits analysés.



L'algorithme tiré des travaux de Pierre Lafon[4] (et fourni par Philippe Cibois) calcule l'indicateur qui fait ressortir **les mots qu'il était très peu probable de trouver *aussi fréquemment (spécificité positive)* et ceux qu'il était aussi peu probable de trouver *aussi rarement* (*spécificité négative)* dans les extraits retenus**.

Tri des occurrences par spécificité

[4] Lafon Pierre. Sur la variabilité de la fréquence des formes dans un corpus. In: Mots, octobre 1980, N°1. pp. 127-165.



Les barres bleues représentent la spécificité positive, la barre rouge la spécificité négative. Les barres représentent l'improbabilité que les mots apparaissent comme ils le font. En clair, plus la barre est large, moins l'apparition du mot dans les conditions observées était probable. Le contenu des cases dans le prolongement des barres renseigne à l'inverse la probabilité en %. Plus elle est basse, moins la forme apparaît de manière prévisible.

- Régler le seuil de spécificité

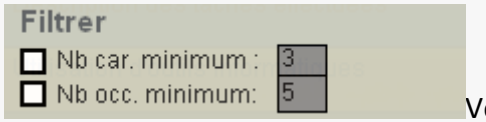
La largeur des barres est fonction du seuil choisi. A droite du bouton d'option permettant de trier les mots par spécificité, se trouve une zone de texte mentionnant le seuil au-delà duquel la probabilité ne doit plus être représentée. Par défaut, le seuil de 10% implique que ne sont représentés que les 10 derniers % d'*improbabilité*. Cela revient à ne voir apparaître de barre bleue ou rouge uniquement pour les probabilités inférieures à 10%. Mettre par exemple un seuil de 5% ne filtrera pas les résultats, mais reviendra à « zoomer » sur les derniers 5%, accroissant ainsi les écarts entre des probabilités proches.

- Afficher (ou non) les spécificités négatives

Par défaut, Sonal n'affiche que les spécificités positives (barres bleues). Cela permet d'éviter un phénomène logique mais problématique qui est que lorsque l'on isole une sous-partie du corpus largement majoritaire, elle ne se caractérise que par les formes absentes, celles-là même qui caractérisent la sous-partie désactivée.

- Filtrer la liste des formes

Vous pouvez choisir de ne retenir que les formes dont la longueur est inférieure à un certain nombre de caractères.

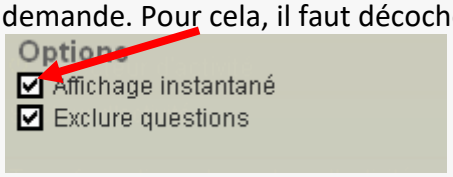


Vous pouvez également choisir un plancher pour le nombre d'occurrences. Celui-ci doit être fonction du nombre d'occurrences total et de la finesse du filtrage.

Dans le cas d'une analyse factorielle, il peut être bon de placer un plancher d'affichage adapté pour limiter le nombre de formes étudiées à moins de 9999.

- Désactiver l'affichage instantané

Pour les corpus d'une taille raisonnable, il est assez pratique que Sonal affiche immédiatement les tris et compte les occurrences à chaque modification du filtrage, mais, dans le cas de corpus très volumineux, cela peut s'avérer plutôt chronophage, surtout si l'on souhaite un filtrage complexe et que Sonal lance un calcul des occurrences à chaque ajout de critère.
C'est pourquoi il est possible de demander que les calculs ne soient effectués qu'à la demande. Pour cela, il faut décocher la case « Affichage instantané » :



Cela implique que pour pouvoir afficher les résultats des calculs, il faudra dorénavant cliquer sur « Rafraîchir » :

- Inclure les questions dans l'analyse lexicale

Par défaut, Sonal fait le choix de considérer que le vocabulaire des questions n'a pas à apparaître dans les analyses. On peut cependant les faire apparaître en décochant la case « exclure questions ». Sonal va alors recharger toute la base lexicale en réintégrant le vocabulaire des interrogateurs.

16.5 Revenir au contexte d'apparition d'une forme

En isolant consécutivement les différentes thématiques, ou les différents (types d') entretiens, on accède ainsi très rapidement aux mots qui leur sont le plus ou le moins significativement liés.

Lorsqu'une forme/un mot vous apparaît comme digne d'intérêt, parce qu'il/elle est spécifique ou pour une autre raison, vous pouvez immédiatement retourner aux extraits qui contiennent ce mot, ou ces mots s'il s'agit d'une forme lemmatisée. Cette fonctionnalité, généralement appelée « concordances » ou KWIC (pour Key Word In Context), est incontournable dans la plupart des logiciels d'analyse lexicométrique. Elle permet d'étudier les différents contextes dans lesquels la forme apparaît, pour assurer un retour interprétatif aux matériaux, afin d'analyser les différents sens conférés à cette forme par les locuteurs.

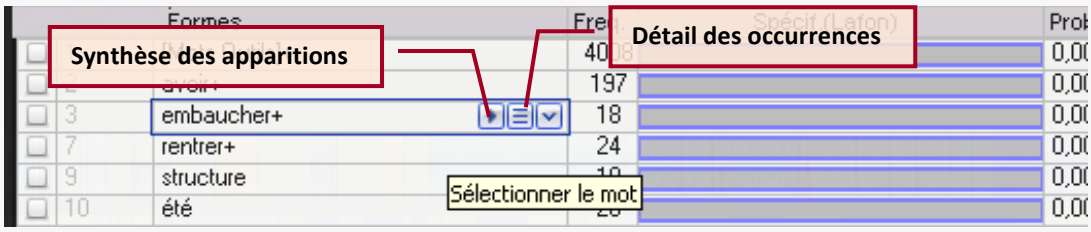
> *Certains logiciels, spécifiquement dédiés à cette tâche, sont bien sûr capables de discerner les contextes, en étudiant les mots associés. Ce n'est pas le cas de Sonal qui, pour l'instant, s'apparente plutôt à un chien de chasse pour gibier d'eau, partant en traque tous azimuts, et qui ramène parfois à son maître des canards en plastique. Une analyse qualitative des résultats quantitatifs est donc indispensable.*

Pour ce faire, deux possibilités s'offrent à vous. Parmi les boutons qui apparaissent à droite du mot survolé, les deux premiers servent à accéder au détail de ses occurrences.

Le premier vous permet d'accéder à la synthèse des extraits contenant le lemme sélectionné (le mot ou les mots regroupés sous un même terme).

La page de **synthèse** s'affiche alors, avec votre sélection.

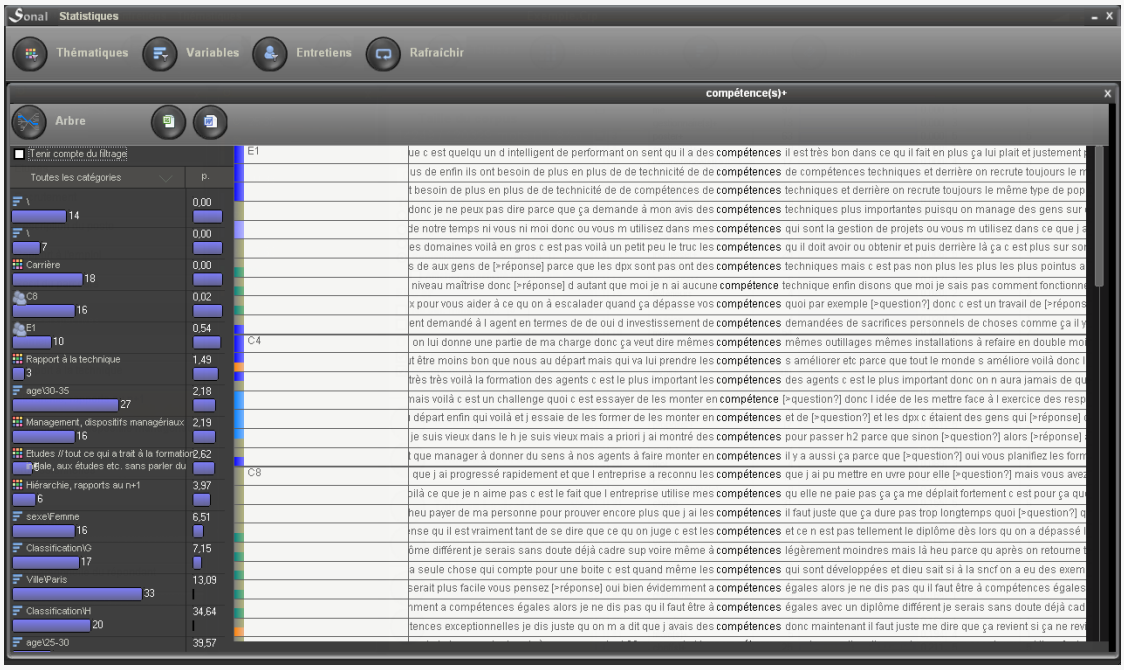
Le bouton représentant une petite liste, situé sur immédiatement à droite du bouton play, vous permet quant à lui d'accéder à une page d'**analyse du mot** qui contient de manière condensée toutes les phrases dans lesquelles le mot recherché apparaît.

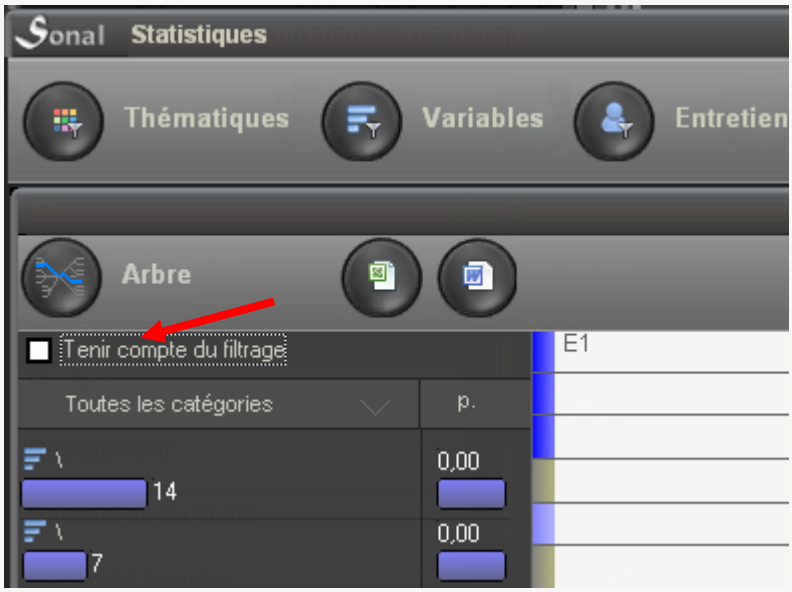


- Analyser un mot en détail

Sur la partie droite de la page, apparaissent toutes les phrases comprenant le mot choisi. Comme ce listing est tiré de la base lexicale, tous les signes de ponctuation et les sauts de page sont supprimés.

Pour les mots très fréquents (au dessus de 500 occurrences) le chargement de cette page peut être assez long : un message vous demandera de confirmer votre commande.

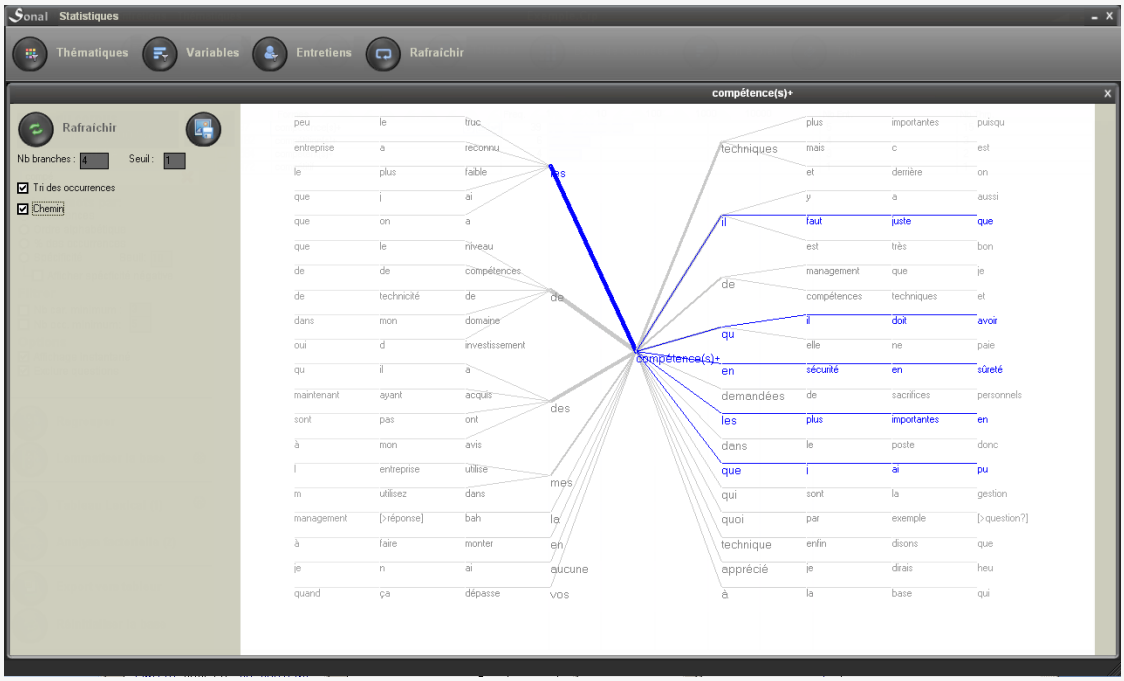




En cochant ou non la case « tenir compte du filtrage », vous choisirez de voir ou non toutes les occurrences du corpus. Le cas échéant, toutes les apparitions du mot choisi seront recensées dans la partie droite de la page, qu'elles soient ou non dans des thématiques ou des extraits filtrés.

Pour faciliter la lecture des contextes, Sonal vous propose une vue sous forme d'arbre, qui permet de trier les contextes avant et après.

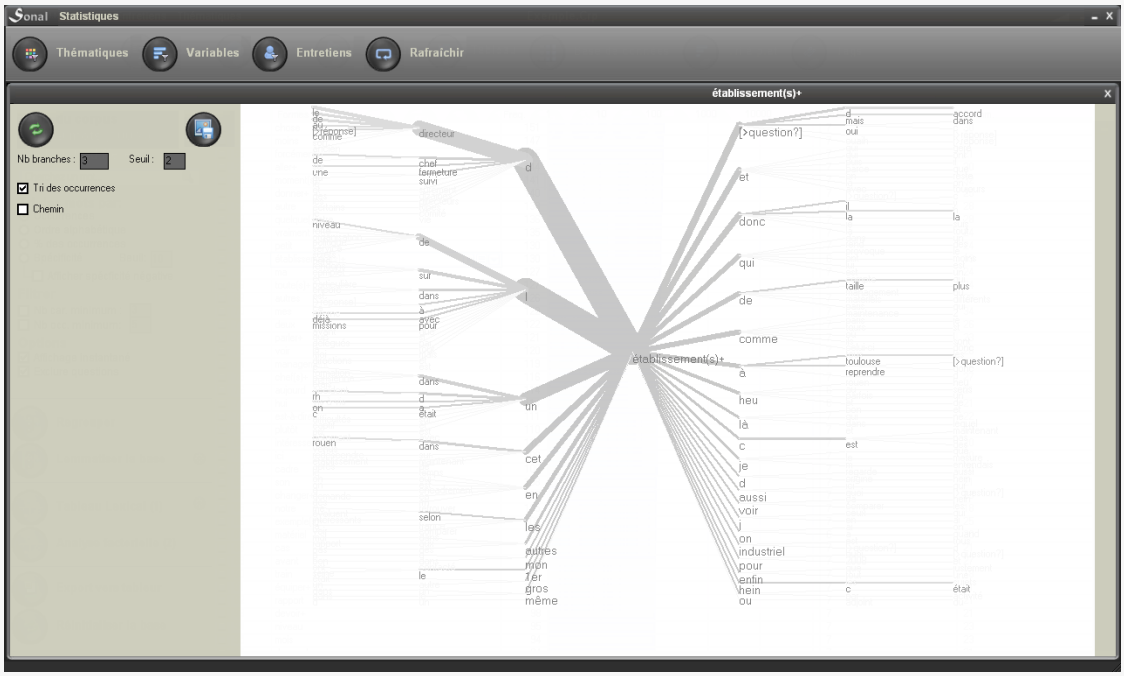
Attention là aussi, la rapidité d'affichage est directement fonction du nombre d'occurrences. Au-delà de 200 ou 300 occurrences, Sonal risque de mettre un certain temps à vous afficher tout l'arbre.



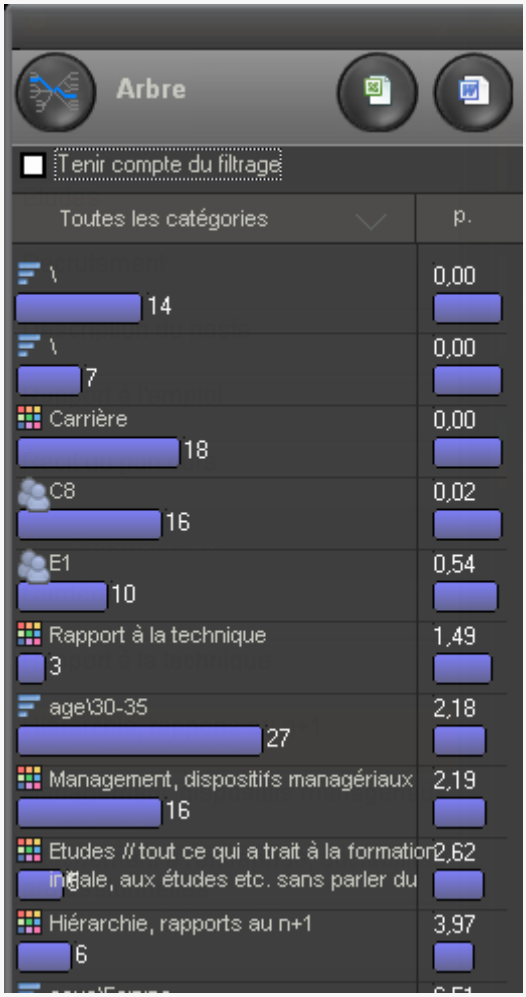
Cliquez sur le bouton « Arbre », pour accéder à cette vue.

Les options sont les suivantes :

- *Nb de branches* : combien de mots avant et après vous voulez voir affichés. Plus vous demandez de branches, plus le calcul est long.
- *Seuil :* le nombre d'occurrences au-delà duquel vous voulez voir apparaître les mots. Cela peut permettre d'y voir plus clair lorsque le nombre d'occurrences est élevé.



- *Tri des occurrences :* cette fonction revient à trier par ordre décroissant les contextes avant et après le mot pivot.
- *Chemin* : cette fonction vous permet de voir, au survol des mots, ce qui précède ou ce qui suit le mot pivot.

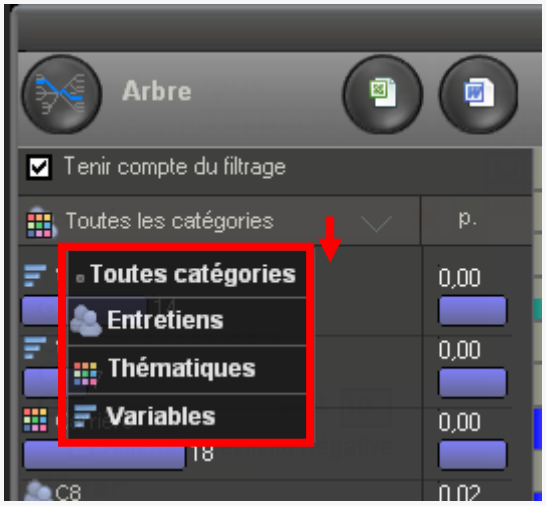


Dans le panneau latéral (à gauche), vous verrez les occurrences cumulées (hors filtrage). Sonal affiche alors toutes les catégories du corpus avec lesquelles la forme est corrélée, indépendamment de la sélection que vous avez faite. On peut dire que vous retournez l'analyse : en général, vous vous penchez sur un mot après avoir isolé une sous-partie du corpus (une thématique et/ou un type de locuteur). L'intérêt de cette fonctionnalité, en l'absence de filtrage, est de vous dire en somme d'*où vous auriez pu arriver*.

Voici le sens des graphiques apparaissant dans la partie latérale gauche de la page :

La barre bleue à gauche représente les effectifs; la barre de droite – bleue ou rouge – représente la probabilité calculée avec l'indice de spécificité de P. Lafon pour que ces effectifs soient observés dans la catégorie comme ils le sont.

Par défaut, vous verrez toutes les catégories en présence mais vous pouvez choisir de ne représenter que certaines catégories d'analyse (thématiques, entretiens, variables…) Cliquez sur l'étiquette ci-dessous pour afficher le menu qui vous permettra de sélectionner les catégories à afficher.



16.6 La lemmatisation

Sur la partie gauche du panneau consacré à la lexicométrie, différents boutons vous permettent de modifier la base lexicale et de lancer des analyses.

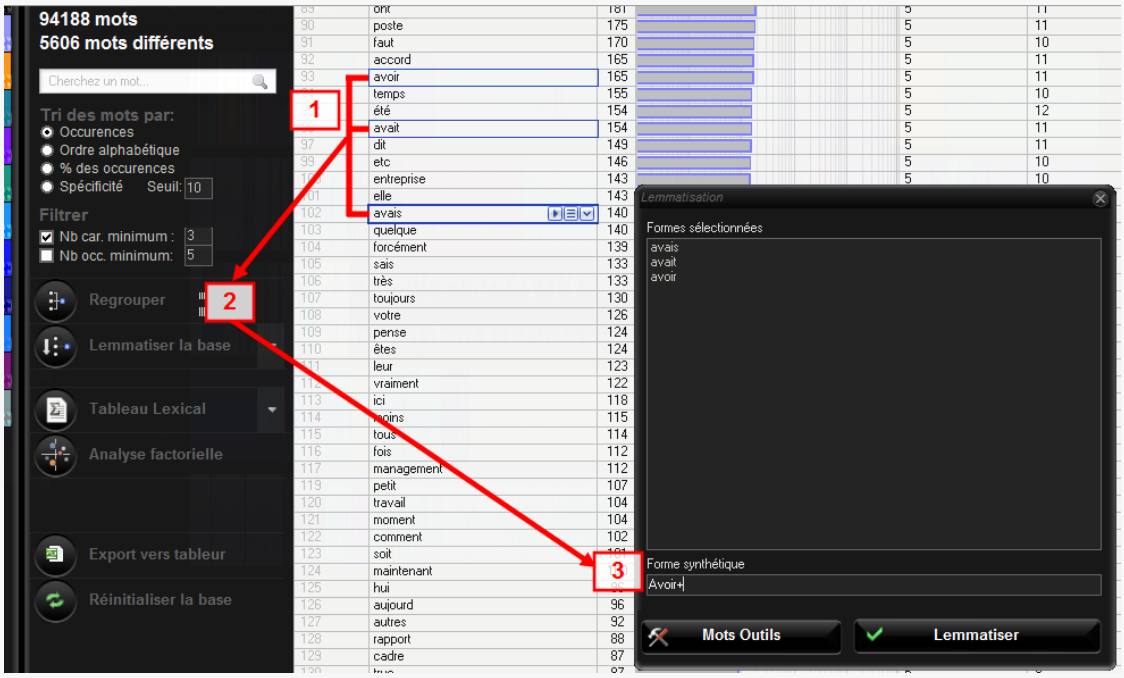
La lemmatisation consiste **à regrouper sous un même terme différents mots** qui partagent par exemple une racine commune, ou qui se réfèrent à un même champ lexical. Concrètement, cela revient par exemple à mettre sous le terme « manger » toutes les formes conjuguées (ou « fléchies ») du verbe manger : (je) mange, (nous) mangeons (qu'il) mangeât etc. Cela peut aussi consister à mettre au singulier tous les pluriels etc. Cela permet de diminuer le nombre d'occurrences différentes dans la base, et de durcir un peu les résultats statistiques.

Sonal vous laisse le choix de lemmatiser ou non votre corpus et de le faire soit de manière automatique, soit manuelle. De plus, pour éviter les écueils liés à l'automatisation, vous pouvez revenir tant que vous le voulez sur les lemmatisations opérées, car le contenu d'un lemme reste accessible et les formes regroupées peuvent toujours faire l'objet d'une analyse individualisée, et être « détachées » du groupe.

- La lemmatisation manuelle : « regrouper »

Le bouton « regrouper » permet de créer des lemmes "manuellement".

Etapes de la lemmatisation manuelle



Vous devez d'abord sélectionner les mots que vous voulez regrouper (1). Pour cela cliquez simplement sur les mots dans la liste. Pour faire une sélection collective, enfoncez le bouton de la souris sur le premier mot de la série puis descendez la souris sans relâcher le bouton. Cliquez ensuite sur le bouton "Regrouper" (2) qui vous affiche une **fenêtre de lemmatisation** récapitulant toutes les formes sélectionnées. Dans la zone de texte en bas de la page, donnez un nom pour le groupe créé (3), (pour faciliter la remémoration qu'il s'agit d'une forme synthétique, vous pouvez prendre le parti de placer un petit "+" à la fin). Cliquez enfin sur "lemmatiser" pour valider. Votre nouveau groupe va être créé. Si le nom de groupe existe déjà, les mots seront simplement ajoutés au groupe existant.

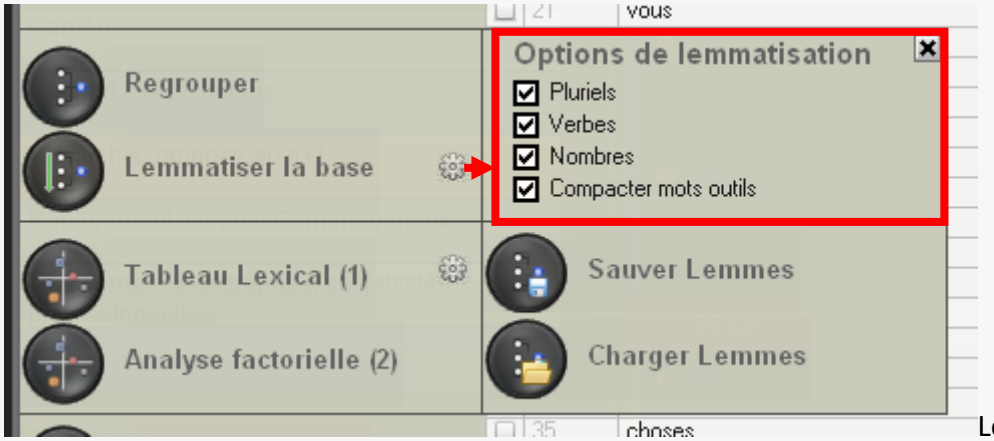
La base est partiellement réécrite puis triée. Cette opération peut prendre un certain temps si le groupe contient beaucoup de mots différents…

- Définir les mots outils

Toutes les formes présentes dans un corpus ne présentent pas forcément un grand intérêt. On ne peut cependant pas présumer de la liste de formes que chacun souhaitera écarter de l'analyse.

Le bouton « mots outils » en bas à gauche de cette page permet d'envoyer les mots sélectionnés dans un lemme spécifique, celui des « mots outils », qui regroupe toutes les formes dont vous n'aurez pas l'usage dans l'analyse. On les appelle parfois des « mots vides » par opposition aux « mots pleins » qui véhiculent du sens. Ces mots outils sont mémorisés dans un fichier texte nommé "Outils.txt", qui est placé dans le répertoire "Dico\" du dossier du programme (généralement C:\Program Files\Sonal\). Cela implique que **votre dictionnaire des mots outils est partagé entre vos différents corpus**. La liste des mots outils est vierge à l'installation, vous devez l'enrichir progressivement. Une bonne méthode est de trier les mots par occurrences avant toute lemmatisation. Les mots les plus courants sont généralement les moins significatifs et peuvent être éliminés collectivement.

- La lemmatisation automatique



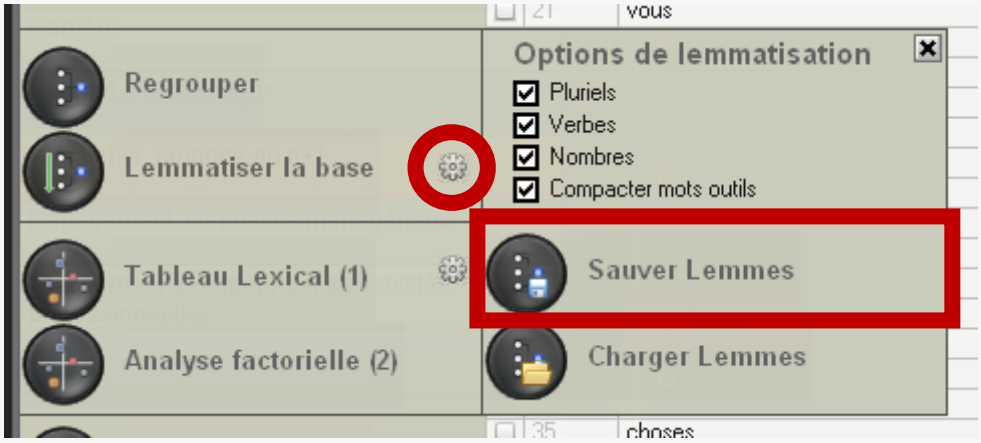
Le bouton

"Lemmatiser la base" propose une lemmatisation automatique de toute la base. En cliquant sur le petit engrenage à droite du bouton, vous pourrez choisir les différentes options. Pour l'instant, est proposée la lemmatisation des verbes, des pluriels (simples ! (s), pour l'instant…), des nombres et des mots outils. A chaque fois qu'un groupe est créé, la forme synthétique affectée à tous les mots qui en font partie est suivie d'un "+".

Cette procédure est très rudimentaire et ne gère aucune exception. Aussi, il est généralement nécessaire de revenir sur les regroupements opérés.

- Enregistrer les lemmes créés

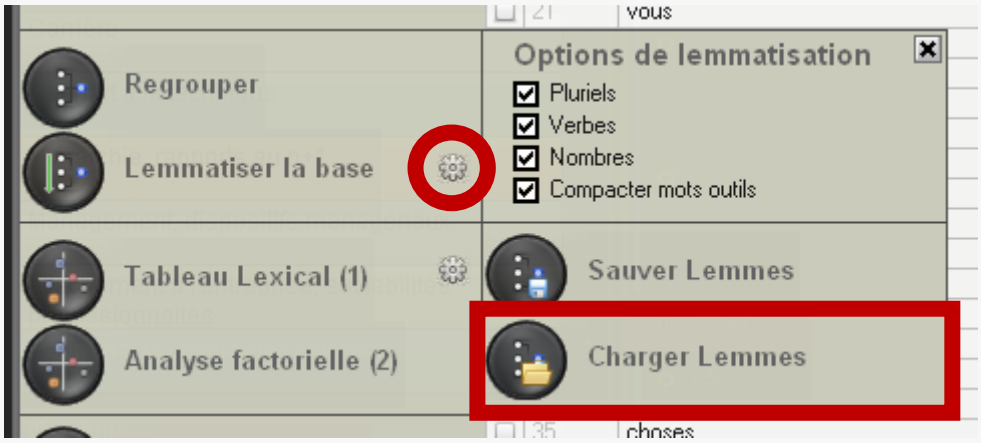
Afin d'éviter de perdre son travail de lemmatisation lorsqu'il est nécessaire de réinitialiser la base, vous pouvez utiliser la fonction sauver lemmes qui apparaît lorsque l'on clique sur la flèche à droite du bouton « lemmatiser la base ».



Sonal crée alors un fichier portant le nom du corpus avec l'extension *.lem
Ce fichier mémorise, pour chaque forme donnée, le lemme associé.

- Recharger les lemmes créés

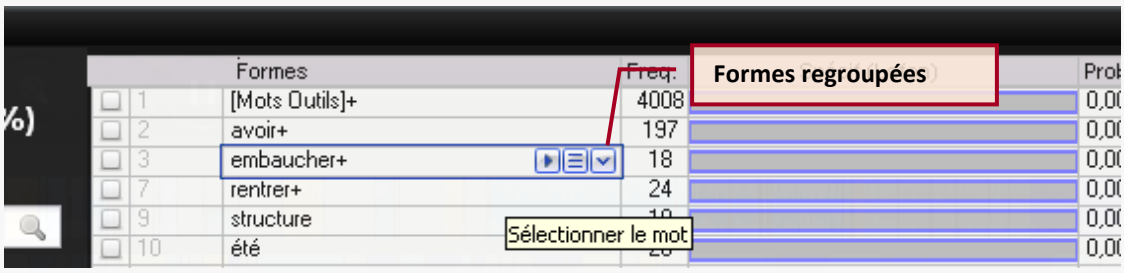
Si la lemmatisation de la précédente base a été sauvegardée, il est possible de recharger les lemmes créés en cliquant sur « charger lemmes » :



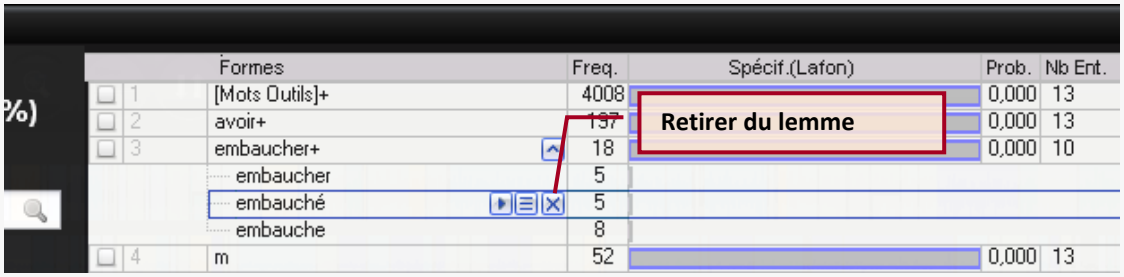
Charger les lemmes revient à parcourir la base lexicale courante et à affecter, à chaque forme, la forme lemmatisée qui était dans la base précédente. Cela peut faire gagner un temps précieux si des groupes de mots très spécifiques avaient été créés précédemment.

16.7 Gérer les groupes créés

Après regroupement, vous avez la possibilité d'ouvrir les groupes de mots créés en cliquant sur la flèche à droite du mot.



Apparaissent alors, sous forme d'arborescence, les différentes formes regroupées sous un lemme commun. Vous pouvez dès lors regarder le nombre d'occurrences de ces différentes formes, et éventuellement "détacher" certains mots du groupe, en cliquant sur la petite croix à droite du nom.



16.8 L'analyse factorielle

Dernière fonction à découvrir dans cette page, l'analyse factorielle du vocabulaire, méthode programmée par Philippe Cibois pour Trideux-Mots, et qui s'inspire des travaux de Lebart et Salem. Pour plus de détails sur cette méthode et sur l'interprétation des résultats produits par Trideux, je vous renvoie à ce chapitre d'ouvrage[5] et à cet article[6].

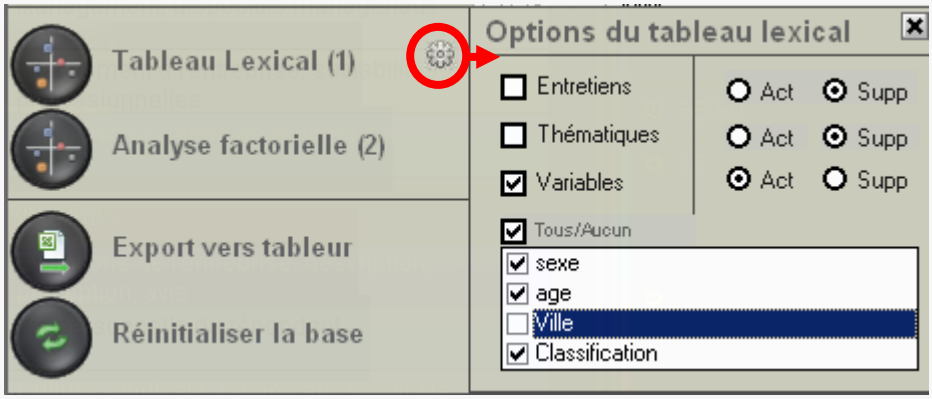
---

[5] Cibois Philippe, Le traitement des mots associés à des questions : l'apport du logiciel Trideux in Didier Demazière, Claire Brossaud, Patrick Trabal et Karl Van Metter (dir.), Analyses textuelles en sociologie. Logiciels, méthodes, usages, Rennes, Presses Universitaires de Rennes, 2006, p.156-173

[6] Cibois Philippe, Eclairer le vocabulaire des questions ouvertes par les questions fermées : le tableau lexical des questions", *Bulletin de méthodologie sociologique*, n°26, 1989, p.12-23.

Dans Sonal, cette méthode procède en deux étapes. Il faut d'abord créer un **tableau lexical,** qui recense, pour chaque mot, le nombre d'apparitions dans les différentes catégories du corpus. Attention, *le tableau lexical est produit à partir des extraits maintenus actifs, et sur les formes retenues après filtrage*. Cela implique qu'une approche globale des matériaux par l'intermédiaire d'une analyse factorielle doit logiquement se faire sur l'intégralité du corpus. Cependant, en cela comme en d'autres choses, Sonal vous laisse une absolue liberté.

En cliquant sur l'engrenage situé à droite du bouton, vous pourrez choisir les catégories à afficher en colonnes dans le tableau lexical, et déterminer si elles joueront un rôle actif ou illustratif (supp.) dans l'analyse factorielle :

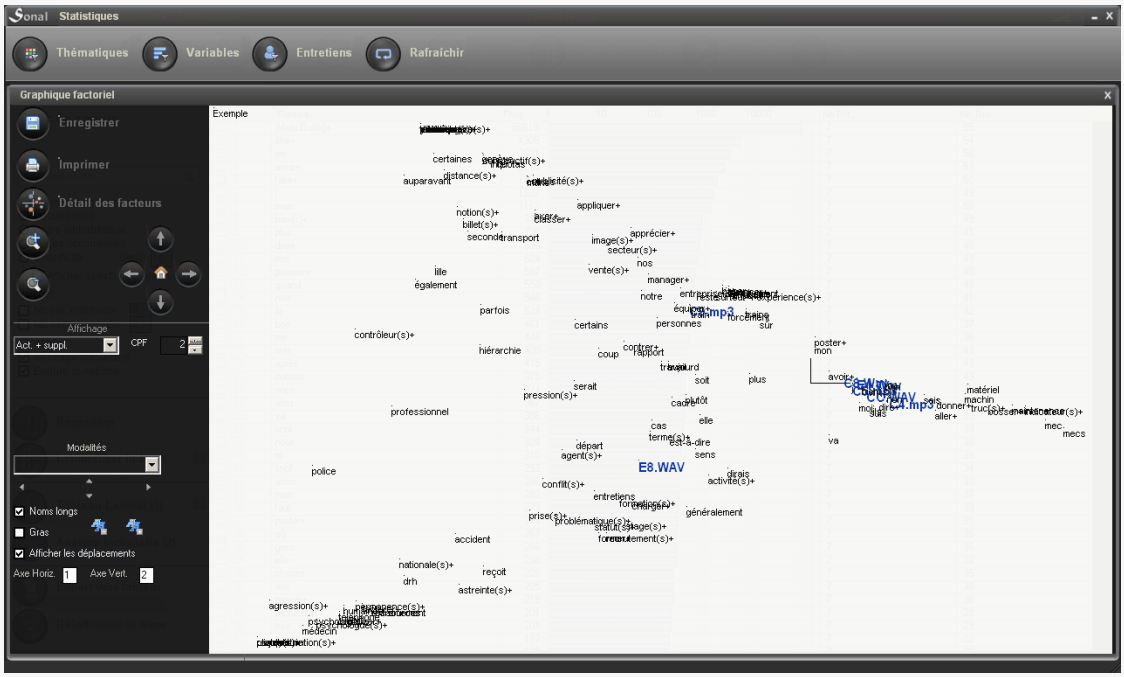


Si vous cochez la case « Variables », vous pourrez choisir quelles variables intégrer dans le tableau lexical.

- ✓ Assurez-vous qu'il n'y a **pas plus de 9999 mots** à intégrer au tableau lexical (utilisez les fonctions de filtrage par occurrence pour
- ✓ Assurez-vous qu'il y a **au moins une catégorie active** parmi les trois disponibles (Entretiens, Thématiques, Variables).
- ✓ Assurez-vous que les catégories analytiques utilisées dans le tableau lexical sont présentes dans votre corpus. Si vous n'avez pas créé de variable par exemple, le fait de demander un décompte des occurrences par variable fera planter le système…

Le tableau est créé. L'affichage du plan factoriel vous est proposé. Pour le voir à nouveau, cliquez sur le bouton "Analyse factorielle".

La fenêtre d'analyse factorielle s'ouvre alors.



Ce plan, que vous pouvez paramétrer avec les différentes options situées sur la partie gauche de l'écran, vous permettra de faire ressortir des typologies articulant le langage utilisé (en noir) et les différentes catégories (en bleu). Pour une meilleure compréhension de la méthode et de la manière dont il faut lire ses résultats, je vous renvoie aux publications de P. Cibois sur le sujet.

17 L'analyse chronométrique

17.1 A quoi sert l'analyse chronométrique?

L'analyse chronométrique vous permet d'étudier les temps de parole consacrés aux différentes thématiques de votre corpus. En analysant le temps passé sur chaque sujet, elle répond à la question de savoir **qui parle de quoi, et pendant combien de temps?** Ces informations peuvent être précieuses pour déceler ou confirmer de grandes tendances dans la propension des différents types de répondants à parler de différents sujets.

- Avertissement préalable : encoder n'est pas prouver

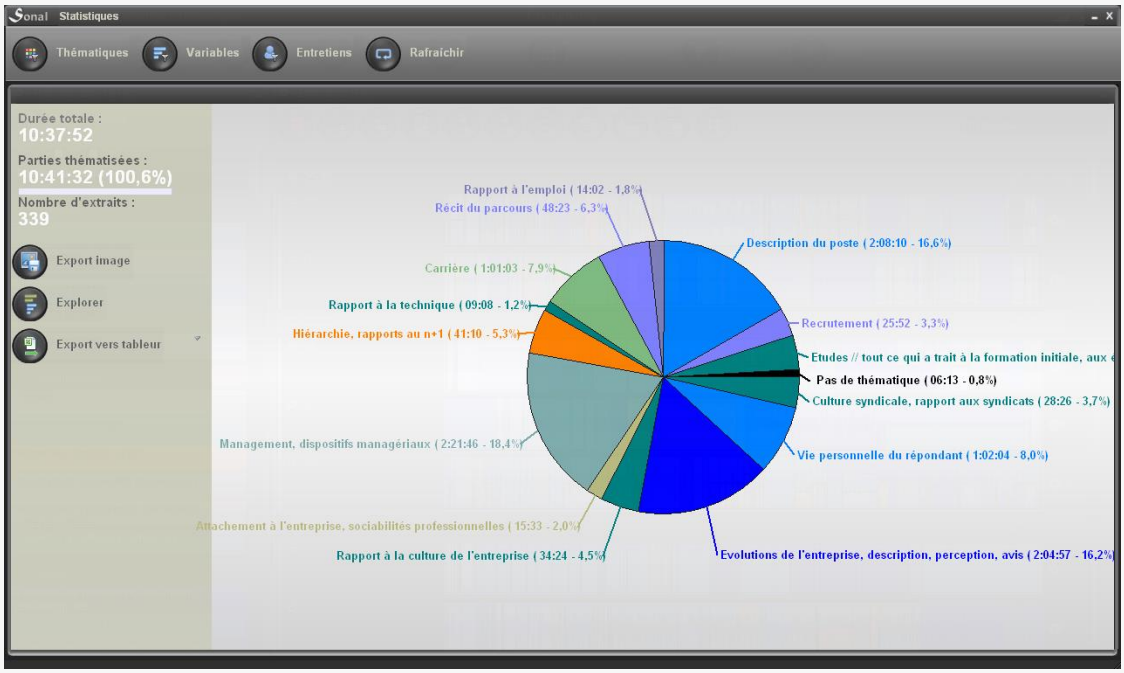
Bien qu'exprimés sous forme de chiffres, les résultats de l'analyse chronométrique ne doivent pas jouir d'une présomption d'objectivité disproportionnée. Le recours à un codage réflexif sert plus à tester des schémas interprétatifs ou à interroger ses propres pratiques, qu'à démontrer quoi que ce soit de manière irréfutable. Il faut bien comprendre que l'encodage thématique, lorsqu'il est l'objet d'une analyse compréhensive basée sur la relecture ou la réécoute des extraits retenus, représente un moindre risque d'abus interprétatif que l'analyse globale des temps de parole qui est proposée ici. Donc, même si la quantification présente une efficacité synthétique séduisante, il faut tenir compte du fait que son objet premier (un temps de parole consacré à un thème) doit être étudié avec mesure :

> *La première question est logiquement celle du rôle joué par l'interrogateur. Selon l'intensité des relances, l'intérêt manifesté auprès de l'enquêté, le répondant se sentira plus ou moins invité à approfondir certaines réponses. Dès lors, le temps passé sur un sujet par un répondant quantifie certains attributs d'une interaction sociale, plus que la propension "naturelle" des répondants à s'épancher sur un sujet donné. Cela n'implique cependant pas que les dispositions à s'exprimer sur un sujet soient réparties de manière homogène dans la population, mais il faut garder un certain recul par rapport aux résultats.*

17.2 Réaliser une analyse chronométrique dans Sonal

Pour accéder à l'analyse chronométrique, il suffit de cliquer sur le bouton "Chrono", disponible dans la page de corpus.

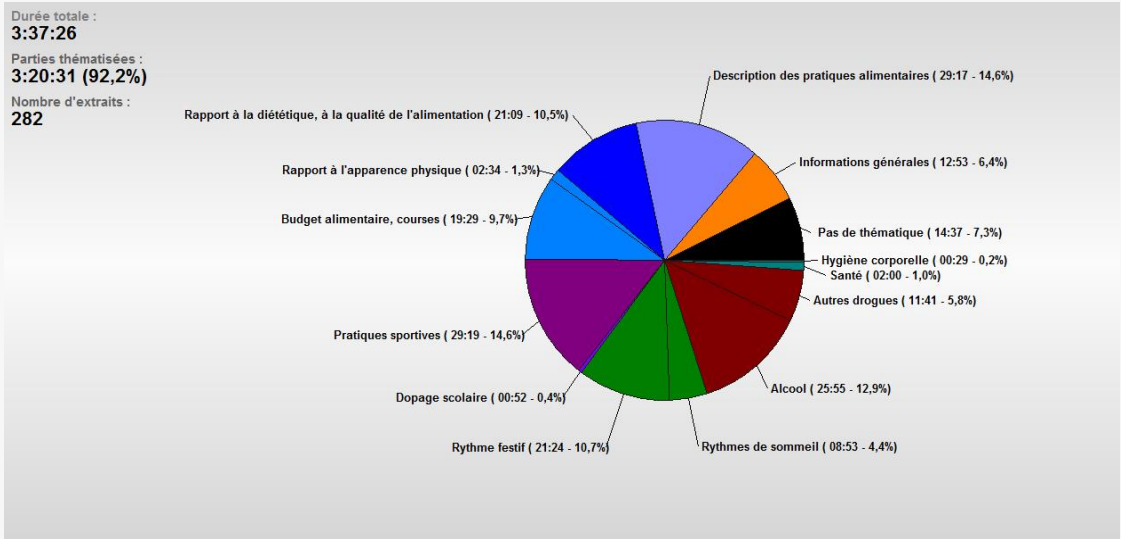
Apparaît alors, sous forme de graphique circulaire, la place des différents sujets dans le corpus.



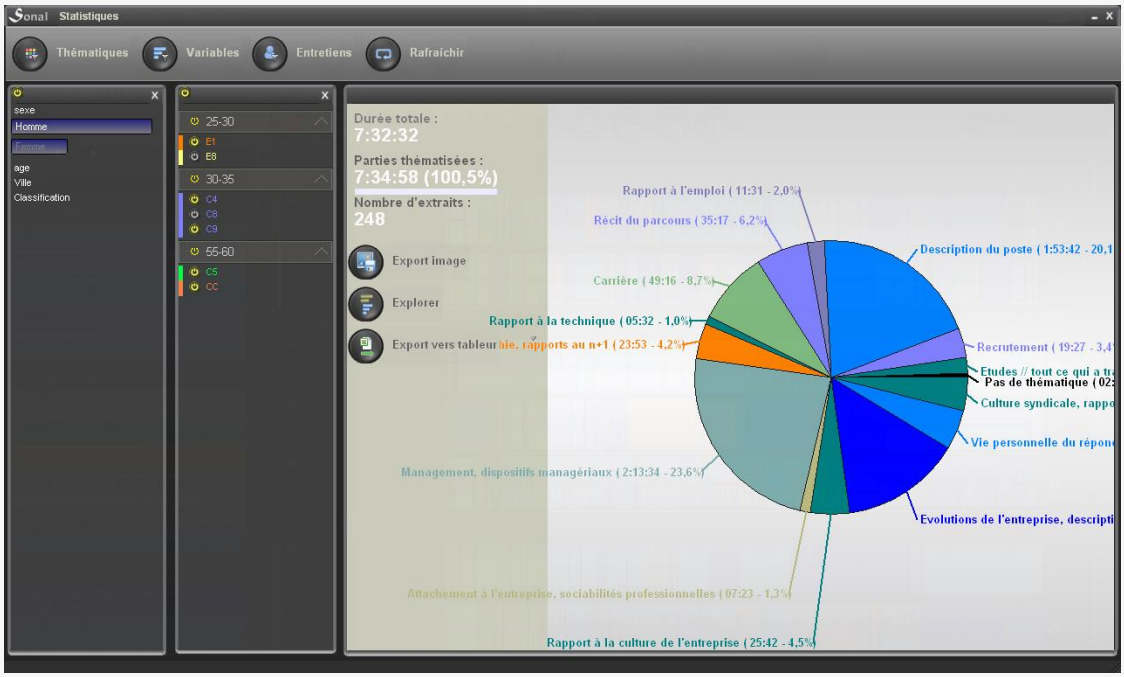
Vous disposez de plusieurs informations :

- le temps total de bande
- la durée totale des extraits et le pourcentage de la bande encodée (celui-ci peut excéder 100% en cas de superpositions d'extraits)
- le nombre d'extraits

Le graphique circulaire peut être exporté en utilisant le bouton "Export image" qui vous permettra d'enregistrer l'image (au format Bitmap) au nom et à l'emplacement voulus. C'est à peu de choses près une copie d'écran.



Pour l'analyse d'un entretien donné, ou d'un type d'entretien, utilisez les boutons de filtrage. Les panneaux (thématiques, entretiens, variables) vous permettront d'affiner vos analyses.



Cette méthode se borne à une analyse locale qui ne permet pas à proprement parler de comparaisons. Vous pouvez évidemment alterner les filtrages pour repérer visuellement les variations entre entretiens ou types d'entretiens, mais ce n'est pas forcément très pratique. Des améliorations de l'interface sont en cours. D'ici là, la meilleure solution pour analyser les temps de parole est encore d'exporter les résultats pour travailler hors de Sonal.

17.3 Approfondir l'analyse : exporter au format *.Xls ou Trideux

Pour aller un peu plus loin, vous pouvez commencer par exporter les données chronologiques dans Excel ou dans n'importe quel tableur, en cliquant sur "Export vers tableur".

NB : L'export tient compte de vos filtrages.

Les renseignements exportés dans le tableur sont les suivants :

- Rappels des entretiens filtrés (éventuellement)
- Durée et pourcentage des différentes thématiques (vous pouvez donc refaire vos propres graphiques hors de Sonal)

Puis sont proposés plusieurs tableaux croisés :

- Place des différentes thématiques dans chaque entretien
- Place des différentes thématiques en fonction des types d'entretiens (autant de tableaux que de variables)

Ces tableaux sont déclinés à chaque fois en secondes, en pourcentages en lignes et en pourcentages en colonnes :

- Le tableau des secondes vous permettra de faire vos propres graphiques à partir des données brutes,
- le tableau des pourcentages en lignes vous permettra de répondre à la question de savoir qui consacre quelle part de son temps de parole aux différents sujets, en comparant éventuellement chaque sous-population à la moyenne,
- le tableau des pourcentages en colonnes apporte l'information complémentaire : dans quelles populations se concentrent les interventions sur un sujet donné.

17.4 Travailler des données chronométriques dans Trideux

Pour analyser vos données chronométriques, vous pouvez également opter pour un export vers le logiciel Trideux de Philippe Cibois, qui permet un travail de dépouillement beaucoup plus fin.

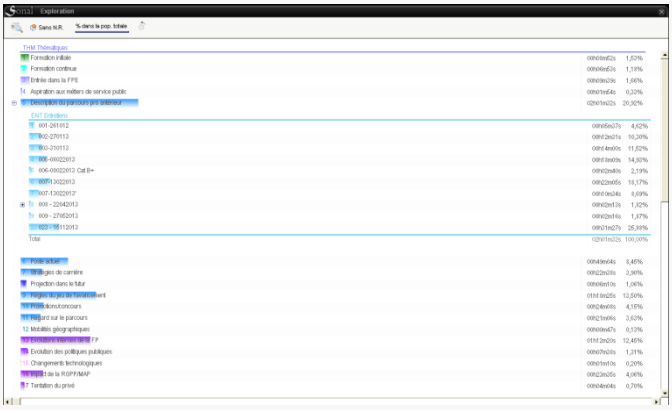
L'export vers Trideux repose sur une astuce : Sonal crée une base de données qui, au lieu de contenir des individus en ligne, contient des unités de temps de parole (des secondes ou des minutes), auxquelles sont conférés les attributs des extraits desquels elles sont issues, c'est à dire à la fois une thématique et des caractéristiques liées à l'entretien.

> *Par exemple, un extrait de 10 secondes sera converti en 10 lignes successives et identiques qui auront toutes les mêmes attributs, ceux de l'extrait et ceux de l'entretien dont il est issu.*

En cliquant ensuite sur le bouton **Export trideux**, vous lancerez l'écriture d'une base lisible par Trideux. Elle sera située dans le dossier du projet, dans un sous dossier nommé "Trideux\". La base créée portera le nom du corpus, suivi de "(HMS)". La base contient pour premières variables les thématiques et les entretiens du corpus. Puis les différentes variables de votre base de données.

Les unités de temps peuvent ensuite faire l'objet d'une analyse similaire à celle qui serait produite à partir de réponses individuelles.

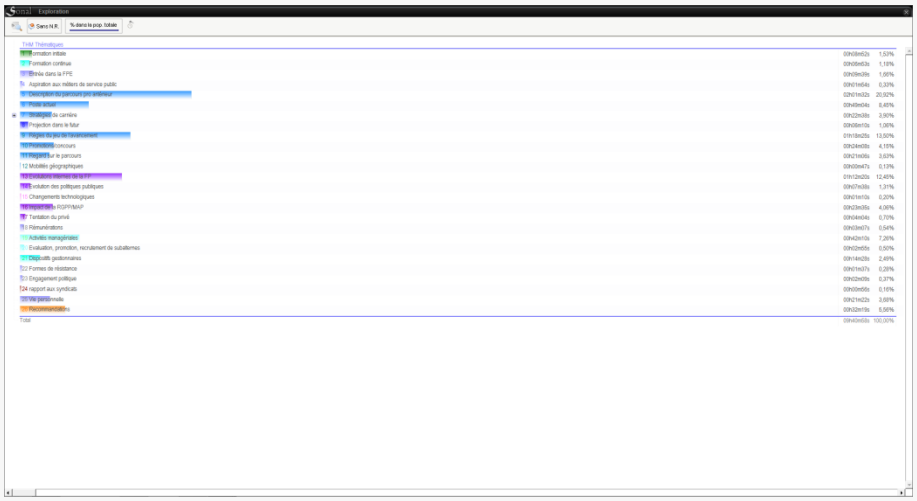
Depuis la version 1.9, Sonal importe une fonctionnalité créée initialement pour Trideux, la fenêtre d'exploration, qui offre une grande souplesse pour étudier les différentes fréquentes du corpus.



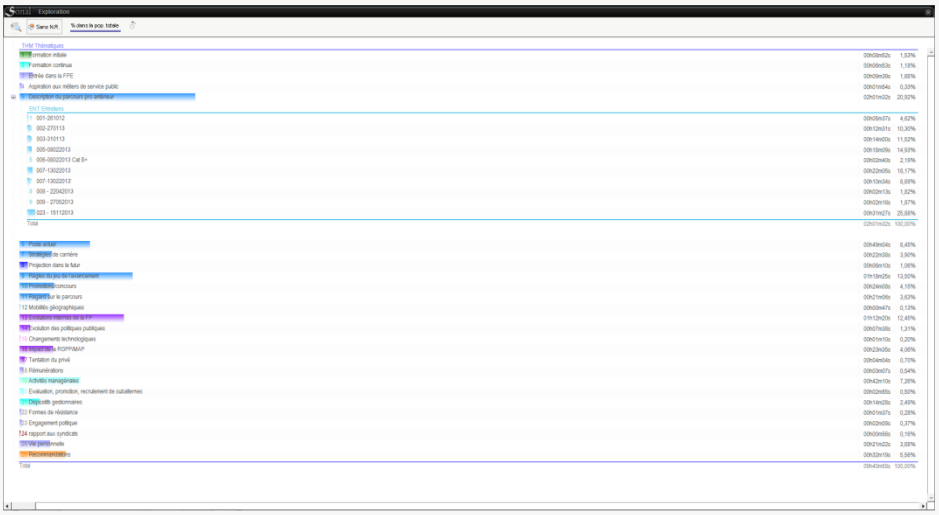
Cette fenêtre, à laquelle on accède en cliquant sur « Explorer », vous permettra de réaliser des tris à plat (analyse de fréquence

univariée) et des tris croisés en allant directement puiser dans la base Trideux créée, sans avoir à ouvrir le logiciel.

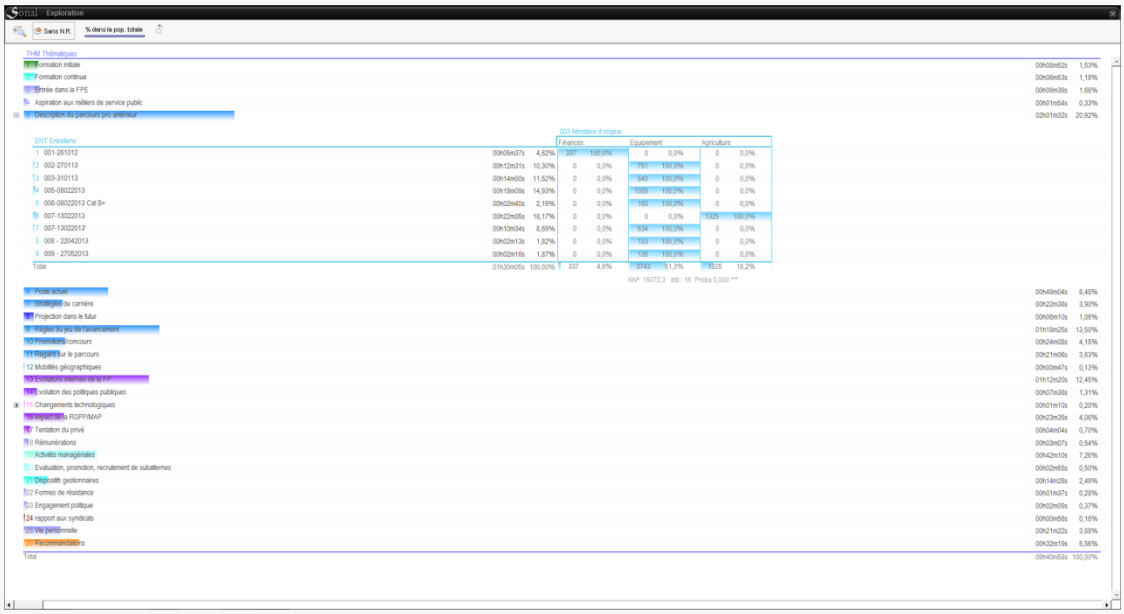
En cliquant sur le bouton qui est situé en haut à gauche de la page, vous aurez accès à la liste des variables de la base de données. En choisissant une variable dans la liste, par exemple "THM - thématiques", vous aurez accès à son tri à plat :



L'intérêt de la fenêtre d'exploration est qu'il vous sera possible de **cliquer sur l'une des modalités pour avoir un nouveau tri à plat**, mais portant uniquement sur la population en question. Vous pouvez par exemple cliquer sur une modalité de votre tri des thématiques et choisir ensuite la variable "Entretiens". Vous aurez alors le temps passés sur ladite thématique dans les différents entretiens. Vous pouvez ainsi « creuser » dans votre base à volonté :



La fenêtre d'exploration vous permet en quelque sorte de faire des « tris croisés à plat ». Mais vous pouvez également y réaliser des tris croisés classiques, en cliquant - au survol d'une variable -, sur le bouton situé en bout de ligne :



Si vous voulez approfondir vos analyses, vous pouvez également installer Trideux et ouvrir votre base chronométrique comme n'importe quelle base.

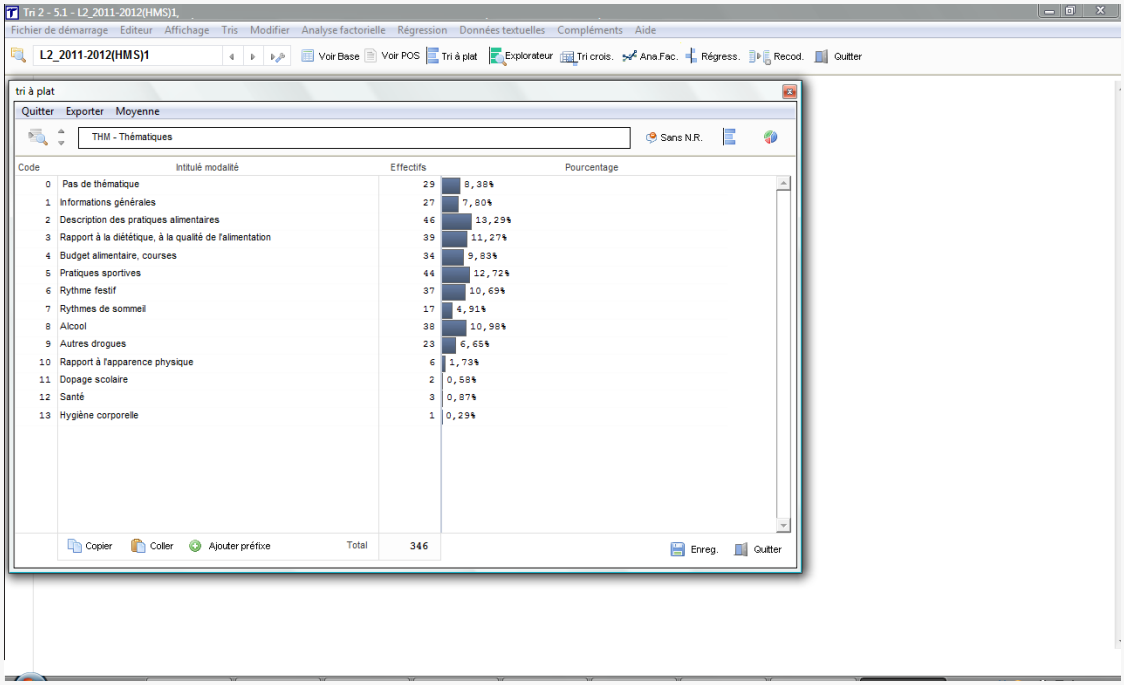
17.5 (se) Lancer (dans) Trideux

> *La version actuellement disponible de Trideux (5.1) ne permet pas encore le démarrage automatique du logiciel après export. Pour l'instant, il faudra lancer Trideux et l'initialiser sur le fichier Tri2.ini qui est situé dans le dossier Sonal. Pour cela, lancez le logiciel, cliquez sur le menu "fichier de démarrage" dans la page principale, puis, sur "Chercher Tri2.ini". Pointez sur le fichier que vous venez de créer, cliquez sur "ouvrir" et vous voilà dans la base.*

Une fois dans Trideux, toutes les fonctionnalités du logiciel vous sont accessibles.

Vous pouvez par exemple accéder, via le bouton Tri à plat, à une représentation en barres des temps de parole consacrés aux différentes thématiques, de la durée des différents entretiens ou le temps total par catégorie de répondant.

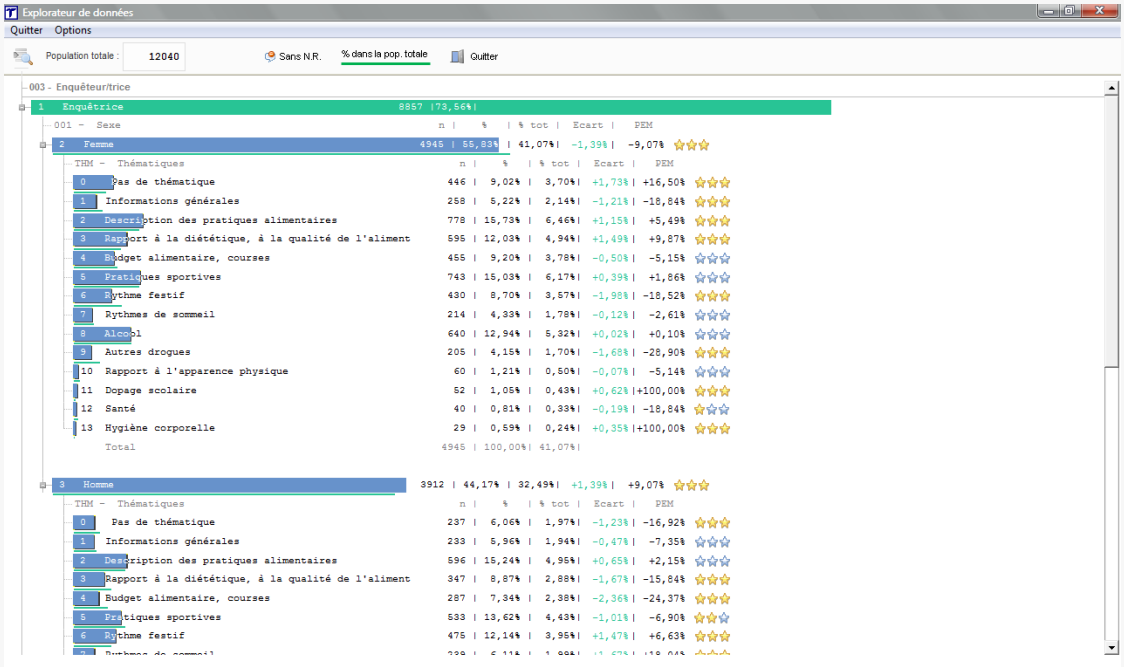
Tri à plat des thématiques (dans Trideux)



Pour étudier plus en détail les résultats de votre tri à plat, vous pouvez utiliser l'explorateur de données, qui vous permettra, de manière intuitive, de "creuser" vos catégories à la manière de poupées russes.

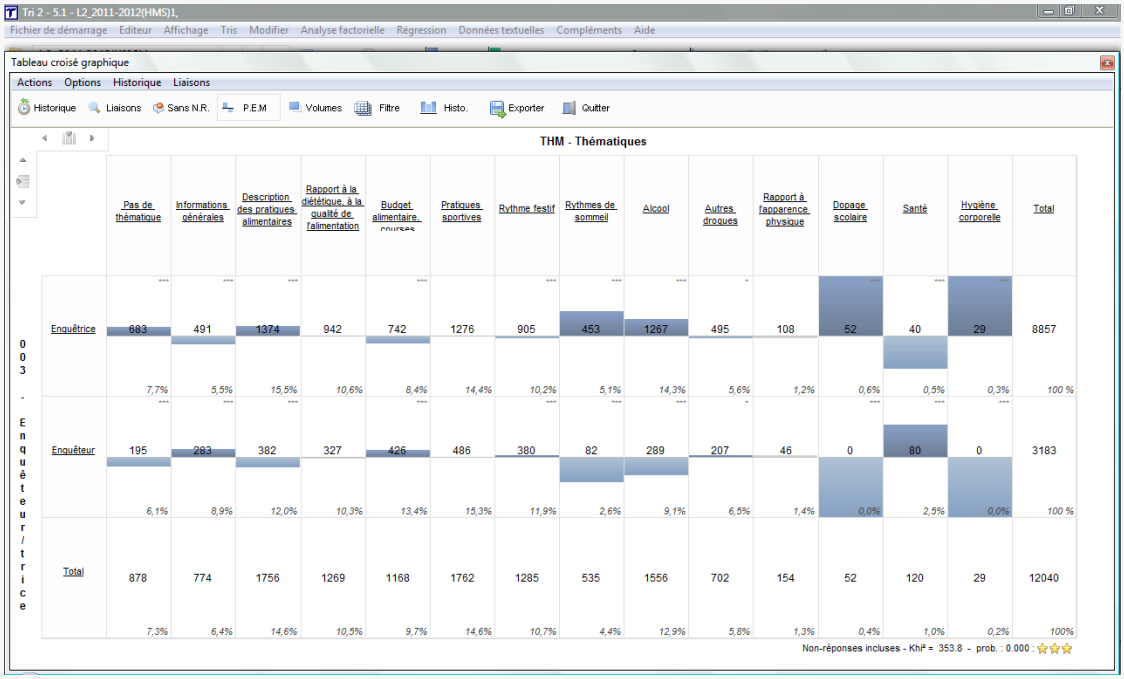
Dans l'exemple ci-dessous, les données sont tirées d'un corpus d'entretiens exploratoires réalisés par des L2 de sociologie de l'Université F. Rabelais, en vue de la réalisation d'une enquête par questionnaire. Les entretiens, relativement courts, ont été encodés thématiquement et annotés, avant d'être redistribués sous forme de corpus audio aux étudiants. En parallèle, chaque entretien s'est vu qualifier par quelques variables (sexe, âge du répondant, sexe de l'enquêteur etc…), ce qui a permis d'avoir une vue globale de l'échantillon, et d'étudier les variations du contenu des entretiens selon les caractéristiques de l'enquêté comme de l'enquêteur. Dans l'exemple ci-dessous, on étudie la place des thématiques dans les entretiens réalisés par des enquêtrices, selon qu'elles interrogeaient des étudiantes ou des étudiants.

Fenêtre d'exploration de données



La fenêtre de tableau croisé graphique vous donnera la possibilité de réaliser ce genre de croisements, mais de manière plus systématique. La représentation graphique des PEM agissant comme un stimulus pour vous permettre de repérer rapidement les cases contenant des écarts à la moyennes intéressants.

Tri croisé graphique (Trideux)



18 L'analyse des échanges

L'analyse quantitative des échanges ne porte pas **sur les extraits mais sur les interventions**, qui sont délimitées par les balises de locuteurs, qui indiquent les tours de parole. Une intervention est donc une portion de texte contenue entre deux balises de locuteurs différentes.
Les fonctionnalités d'analyse des échanges permettent d'étudier la **répartition des questions et des réponses**, ainsi que leurs **longueurs** respectives (en mots), dans les différentes parties du corpus. On pourra ainsi remarquer si certains thèmes suscitent des réponses plus longues que la moyenne, si l'enquêteur intervient plus souvent avec tel ou tel type de répondant ou sur tel ou tel sujet etc.
L'analyse des échanges est adossée à l'analyse des débits, qui repose sur le dénombrement des mots contenus dans une portion temporelle donnée.

A savoir: afin d'éviter que les informations ajoutées en aparté ("rires" etc.) ne nuisent au décompte des mots, Sonal exclu du décompte tout ce qui est entre parenthèses.

L'analyse globale des échanges se lance depuis la page de corpus, en cliquant sur le bouton « échanges » :

18.1 Le tableau des interventions

Au chargement de la page d'analyse des échanges, tous les extraits du corpus sont découpés en prises de parole qui sont chargées dans le tableau des interventions.
On peut schématiser la structure du tableau des interventions de cette manière :

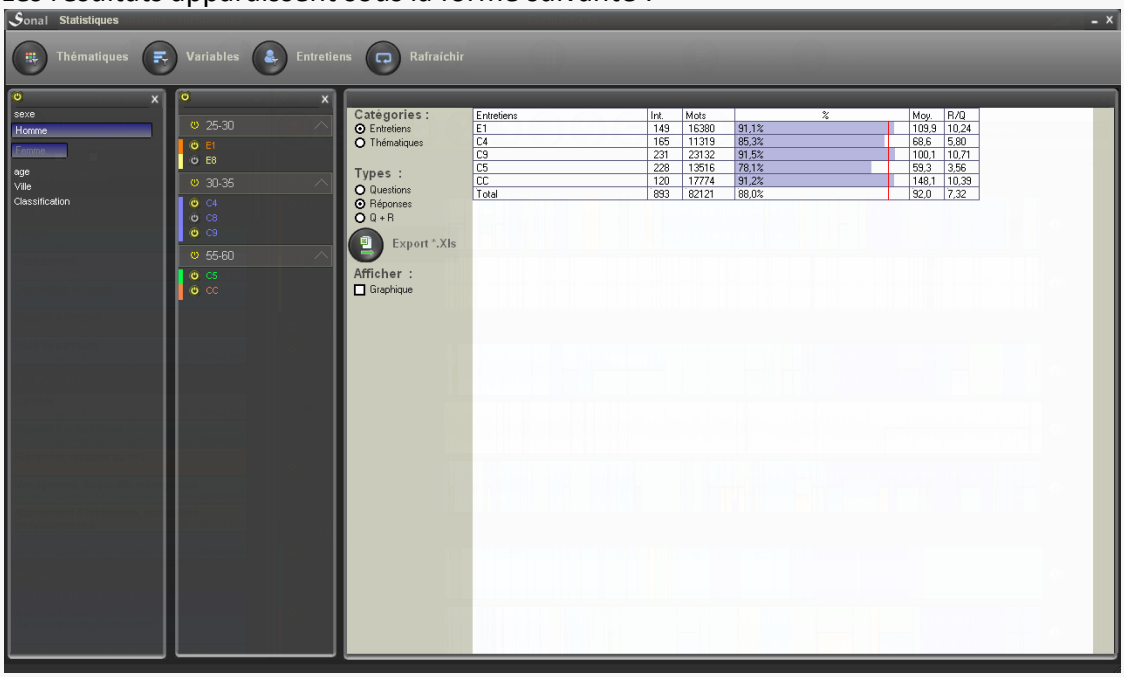
Entretiens	Extraits	Interventions	Caractéristiques de l'entretien			Thématiques de l'extrait			Caractéristiques de l'intervention	
			Sexe	Age	...	1	2	3	Locuteur	Question
Entretien 1	Extrait 1	Intervention 1	H	40-45	...	Oui	Non	Non	X	Oui
		Intervention 2	H	40-45	...	Oui	Non	Non	Y	Non
		Intervention 3	H	40-45	...	Oui	Non	Non	X	Oui
		Intervention 4	H	40-45	...	Oui	Non	Non	Y	Non
	Extrait 2	Intervention 1	H	40-45	...	Non	Oui	Non	X	Oui
		Intervention 2	H	40-45	...	Non	Oui	Non	Y	Non
	Extrait 3	Intervention 1	H	40-45	...	Non	Non	Oui	X	Oui
		Intervention 2	H	40-45	...	Non	Non	Oui	Y	Non
		Intervention 3	H	40-45	...	Non	Non	Oui	X	Oui
		Intervention 4	H	40-45	...	Non	Non	Oui	Y	Non
	...	...	H	40-45	...	...	...	...	...	...

Chaque intervention est contenue dans un extrait qui est lui-même contenu dans un entretien. **Elle récupère donc la thématique et les mots-clés de son extrait** et les **attributs de l'entretien** duquel il est tiré.

On peut ensuite opérer toutes sortes de dénombrements en croisant les caractéristiques du filtrage, exactement comme dans les analyses lexico ou chronométriques. Les mêmes panneaux latéraux sont disponibles pour cumuler les critères.

On peut choisir de focaliser l'analyse sur les entretiens ou sur les thématiques. Les deux ne fournissent pas les mêmes informations et il peut être utile de croiser les approches. Pour savoir sur quels sujets un répondant est le plus prolixe par exemple, on pourra choisir de visionner les résultats par thématique, et filtrer par entretien, à l'aide du panneau latéral.

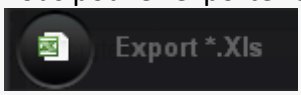
Les résultats apparaissent sous la forme suivante :



Les informations suivantes sont disponibles

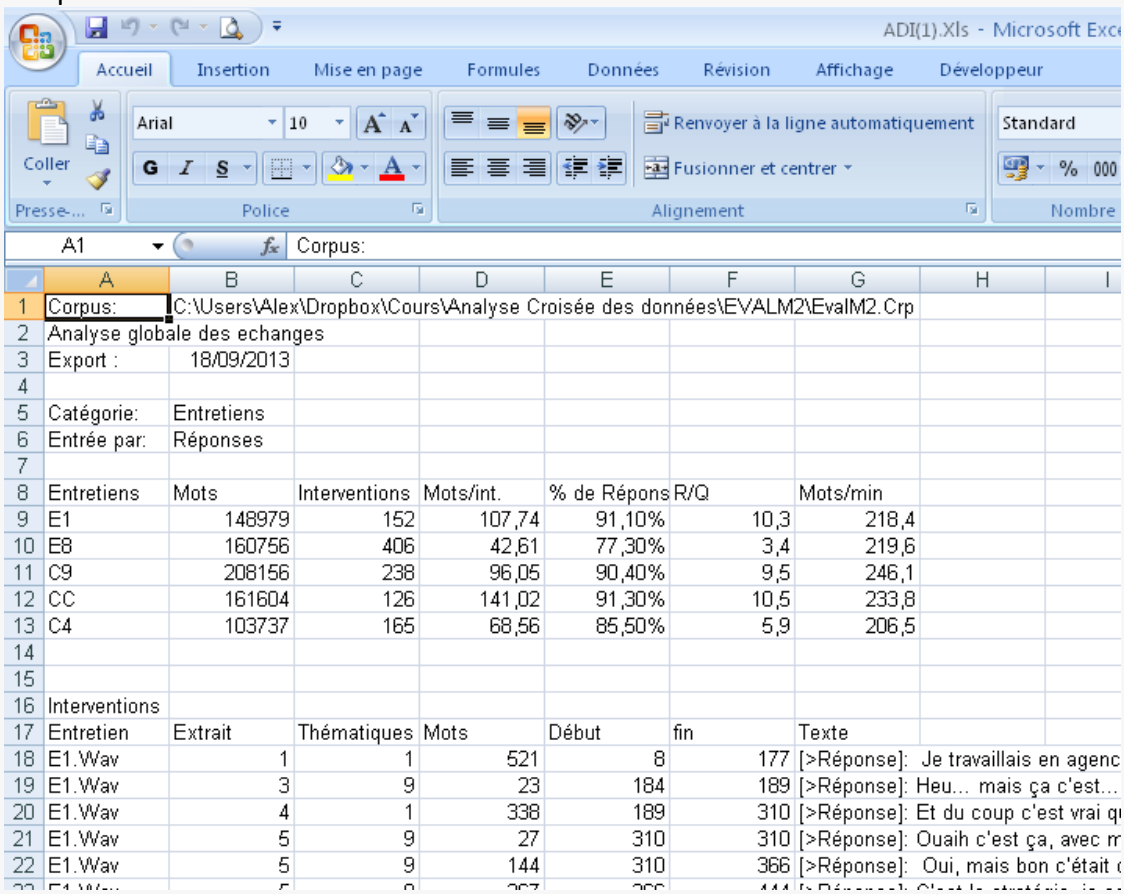
- Int. = nombre d'interventions
- Mots = nombre de mots dans les interventions
- % = Taux de relance. Le taux de relance indique la part relative des questions ou des réponses, selon l'entrée choisie, dans le total des interventions (en mots).
- Moy. = longueur moyenne des interventions
- R/Q = nombre de mots des réponses/nombre de mots des questions. Le ratio R/Q renseigne sur le nombre de mots de réponse par mot de question. Un ratio de trois signifiera que le répondant parle 3 fois plus que l'interrogateur. Ce ratio varie logiquement en proportion du taux de relance.

Vous pouvez exporter ces résultats vers un tableur en cliquant sur le bouton :



Sonal va créer un fichier nommé ADI.xls (pour **A**nalyse **D**es **I**nterventions) dans le répertoire du projet.

Vous aurez à la fois les résultats globaux et le détail des interventions prises en compte.



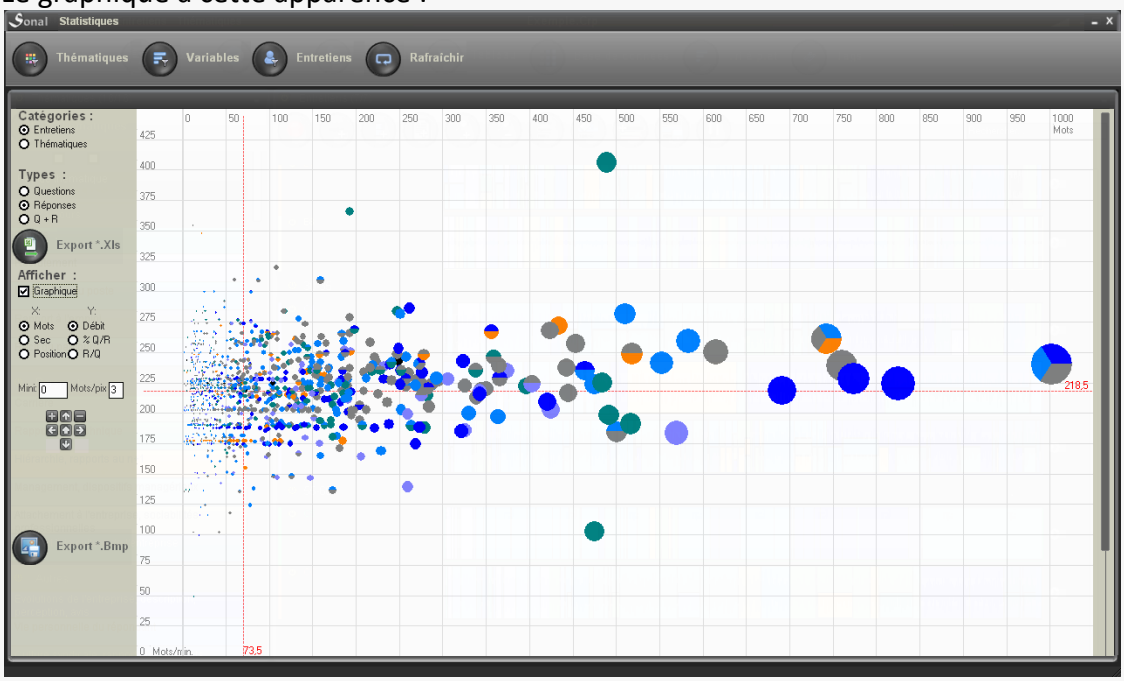
	A	B	C	D	E	F	G
1	Corpus:	C:\Users\Alex\Dropbox\Cours\Analyse Croisée des données\EVALM2\EvalM2.Crp					
2	Analyse globale des echanges						
3	Export :	18/09/2013					
4							
5	Catégorie:	Entretiens					
6	Entrée par:	Réponses					
7							
8	Entretiens	Mots	Interventions	Mots/int.	% de Répons	R/Q	Mots/min
9	E1	148979	152	107,74	91,10%	10,3	218,4
10	E8	160756	406	42,61	77,30%	3,4	219,6
11	C9	208156	238	96,05	90,40%	9,5	246,1
12	CC	161604	126	141,02	91,30%	10,5	233,8
13	C4	103737	165	68,56	85,50%	5,9	206,5
14							
15							
16	Interventions						
17	Entretien	Extrait	Thématiques	Mots	Début	fin	Texte
18	E1.Wav	1	1	521	8	177	[>Réponse]: Je travaillais en agenc
19	E1.Wav	3	9	23	184	189	[>Réponse]: Heu... mais ça c'est...
20	E1.Wav	4	1	338	189	310	[>Réponse]: Et du coup c'est vrai q
21	E1.Wav	5	9	27	310	310	[>Réponse]: Ouaih c'est ça, avec m
22	E1.Wav	5	9	144	310	366	[>Réponse]: Oui, mais bon c'était

18.2 l'analyse graphique des interventions

Sonal propose un outil de visualisation des interventions qui rend possible l'analyse graphique des échanges. Parmi les dimensions utilisées pour la réalisation du graphique, le débit en mot/minutes des interventions, qui impose un calcul des débits au premier chargement.

On obtient une représentation graphique des interventions en cliquant sur la case « graphique » dans la colonne de gauche.

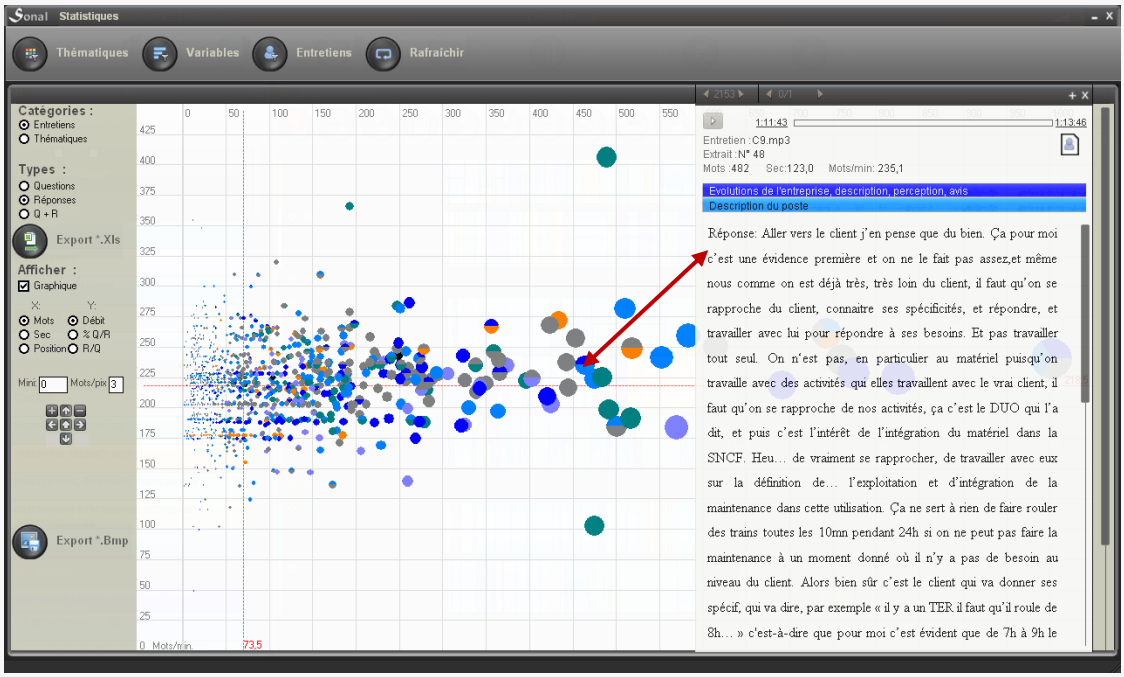
Le graphique a cette apparence :



Chaque cercle représente une intervention. La taille du cercle est proportionnelle au nombre de mots.

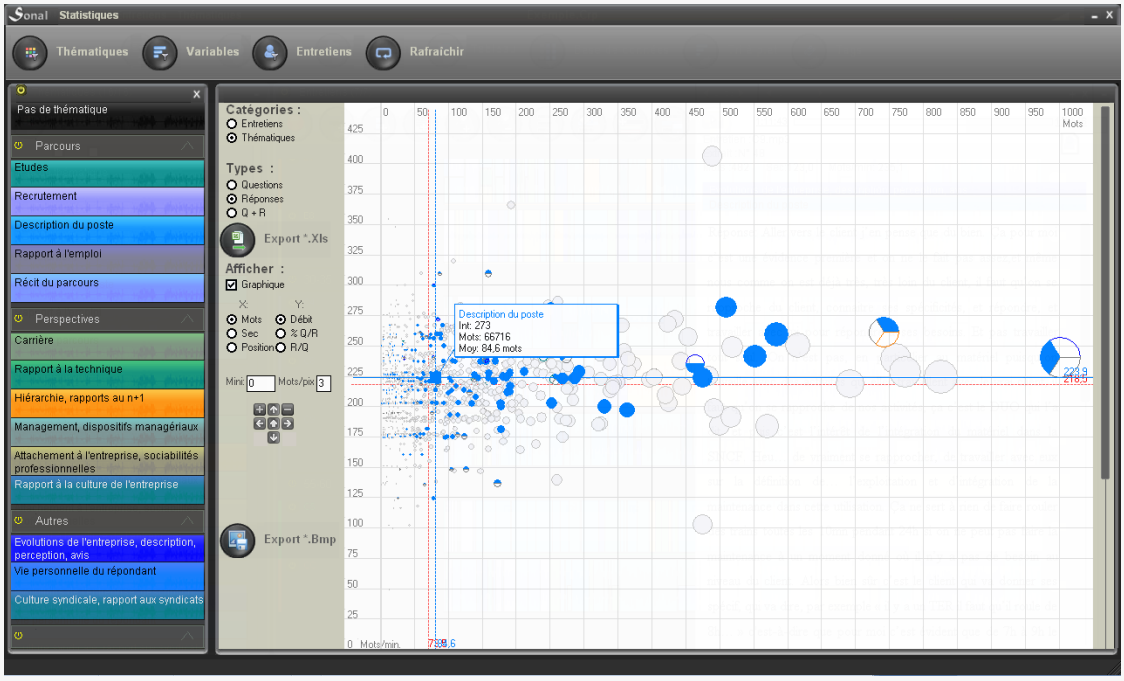
En abscisse, vous pouvez faire figurer la longueur en mots de l'extrait, sa longueur en secondes, ou la position chronométrique de l'intervention. En ordonnées, le débit, le % de questions ou de réponses, ou le ratio R/Q.

Ce graphique vous permet de repérer les interventions les plus longues et/ou les plus rapides. **En cliquant sur une bulle, vous pourrez relire et/ou réécouter le contenu de l'intervention.**



L'approche graphique peut vous permettre de visualiser des interventions atypiques. Soit qu'elles soient très longues, ce qui implique que l'interrogateur s'est abstenu de couper le locuteur, soit très rapides ou très lentes, ce qui peut induire une forte implication émotionnelle du répondant, ou une grande prudence.

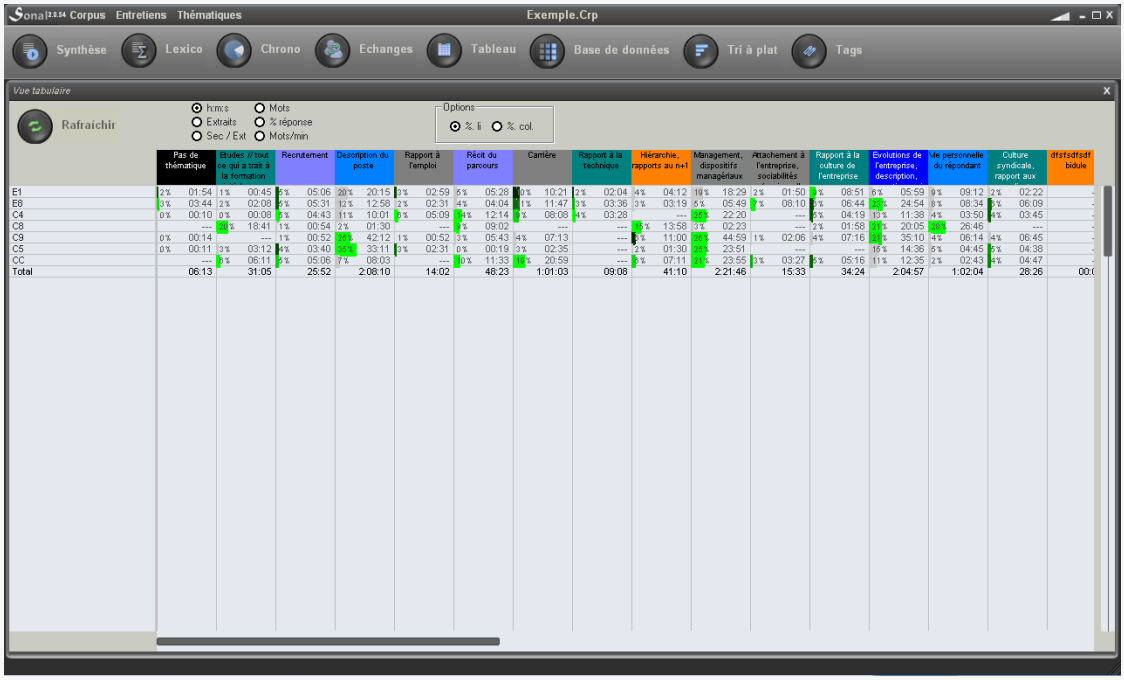
Vous pouvez enfin utiliser cette interface pour visualiser les écarts entre les longueurs et débits moyens dans le corpus et dans l'entretien ou la thématique considérée. Pour cela, survolez à la souris les différents items de la catégorie retenue. *Si vous utilisez les entretiens comme catégorie, seul le panneau latéral des entretiens réagira au survol. Idem pour les thématiques.*



19 L'analyse tabulaire

Bien qu'il puisse sembler exagéré de parler d'analyse s'agissant de la vue tabulaire, l'orientation de cette interface est clairement analytique. Il s'agit de voir le corpus sous forme de tableau.

.



La vue croise les entretiens en ligne et les thématiques en colonnes.

A l'intersection des deux, soit :

- les durées
- le nombre d'extraits
- le nombre moyen de secondes par extrait
- le nombre de mots
- le pourcentage de réponses
- le débit moyen

En cliquant sur les cases, vous accédez à la synthèse des extraits considérés.

En cliquant sur les noms d'entretiens, vous triez les thématiques par % en ligne décroissant.

En cliquant sur les thématiques, vous triez les entretiens par % en colonne décroissant.

Annexes

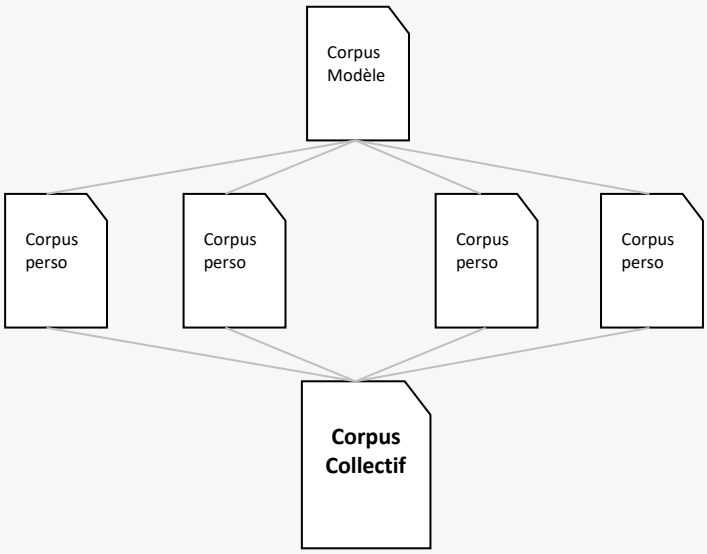
20 Utiliser Sonal dans une recherche collective

Sonal peut vous permettre de construire assez simplement un corpus partagé entre plusieurs chercheurs. Une bonne organisation, et une connaissance minimale de l'organisation du logiciel vous permettront de vous en sortir très facilement.

Le point essentiel à comprendre est que Sonal interprète le contenu des fichiers d'encodage associés aux différents entretiens (*.Rtr) à partir des catégories renseignées dans le fichier corpus (*.Crp).

> *Par exemple, un fichier d'encodage ne contiendra pas le nom des thématiques mais simplement leur numéro (« 001 », « 002 »), c'est dans le fichier Corpus que sera trouvée l'information manquante, à savoir que le code « 001 » ou « 002 » correspond à telle ou telle thématique. De même pour les variables. Chaque entretien contient des codes, mais le dictionnaire des variables est dans le fichier corpus.*

Cela implique simplement que tous les fichiers encodés à partir d'un même corpus partageront une même grille d'encodage (thématique et catégorielle) et pourront être lus identiquement depuis ce même corpus. Autrement dit, pour pouvoir réunir des entretiens encodés par différentes personnes, **il faut que chacun travaille à partir d'une version identique du corpus de départ**, que chacun ajoute ses entretiens dans une version personnelle de ce même fichier de projet et, après les avoir encodés, les réimporte dans un corpus collectif – éventuellement partagé en ligne.



20.1 Créer un corpus modèle

La première chose à faire est donc de créer un fichier corpus aussi complet que possible. Plus votre corpus d'origine prévoira de catégories thématiques et de variables, plus simple sera la fusion ultérieure.

> *En effet, si chacun ajoute de son côté des thématiques, des variables ou des modalités dans sa version du corpus, Sonal créera des codes afférents qui ne seront plus lisibles depuis le corpus partagé (il risque d'y avoir plusieurs catégories thématiques ou modalités différentes pour un même index, par exemple). Bien que cela puisse se corriger assez simplement, cela peut compliquer l'importation des entretiens encodés*

- Construisez la grille thématique

Les instructions relatives à la construction de la grille thématique peuvent être trouvées ici. Il est bon de prévoir autant de thématiques que possible en amont de l'encodage (en se basant sur le guide d'entretien). Au besoin, une fois le corpus fusionné, vous pourrez toujours recoder les thématiques trop peu utilisées.

Cela dit, on ne peut jamais totalement anticiper la liste exacte des thématiques dont on aura besoin au long du processus d'encodage. Une bonne chose à faire est donc de **se laisser un peu de souplesse**. Pour cela, vous avez plusieurs possibilités :

- la plus simple est d'ajouter à la grille des thématiques une catégorie « attrape tout », du type « Pas de thématique adaptée ». Si chaque chercheur y affecte les extraits dont il ne sait pas quoi faire, il sera possible, une fois le corpus fusionné, d'inventorier les extraits non affectés et de les redistribuer dans les thématiques idoines créées *a posteriori*.
- Une autre méthode, un peu plus compliquée, consiste à laisser plusieurs thématiques vierges à chaque chercheur, afin de laisser chacun compléter la grille à sa guise. Attention, cela implique d'anticiper les superpositions de code : il faudra donc affecter des codes spécifiques à chacun. Par exemple, imaginons que vous ayez 25 thématiques au départ. Vous pouvez ajouter 5 thématiques vierges par personne, en affectant les codes 30 à 34 au premier chercheur, 35 à 39 au suivant, 40 à 44 au troisième, et ainsi de suite. Ainsi, chacun pourra utiliser les thématiques vierges qui lui ont été affectées pour créer de nouveaux codes, en se contentant de modifier « ses » catégories thématiques. Par souci de clarté, vous pouvez faire en sorte que chacun ne voie que ses propres thématiques vierges au moment de la distribution des corpus modèles (voir plus bas).

- Renseignez le dictionnaire des variables

La base de données est la partie la plus délicate du partage des corpus. Le dictionnaire des variables est rempli automatiquement au moment de la saisie des valeurs dans la base de données. **Leur ordre d'apparition détermine leur code**. C'est très pratique pour un utilisateur seul, mais cela complique les choses si les plusieurs utilisateurs ajoutent de nouvelles modalités chacun de leur côté. Au moment de la fusion, les dictionnaires ne correspondront pas et ce sera problématique.
La meilleure méthode pour contourner ce problème est logiquement de « **fermer » autant que possible les modalités en amont de la saisie**. Cela veut dire que, comme dans un questionnaire, il va falloir envisager les différentes réponses possibles pour que chacun puisse renseigner les variables en se basant sur une liste existante.
Pour remplir le dictionnaire, vous pouvez ajouter un premier entretien dans votre corpus modèle et saisir successivement (et dans l'ordre) chaque réponse pour l'entretien donné afin de créer les entrées nécessaires dans le dictionnaire des variables. Cela dit c'est un peu fastidieux. Le mieux est encore d'ouvrir directement le fichier *Crp (avec le bloc note) et d'y ajouter manuellement les variables et les modalités.

Les variables sont déclarées entre les balises POS et les modalités entre les balises |DIC|. Voici par exemple à quoi ressemble la déclaration de la variable Sexe dans le fichier CRP, avec les modalités nécessaires.

```
<|POS|
1,Sexe du répondant
>|POS|
<|DIC|
1,1,Homme
1,2,Femme
1,3,Inconnu
>|DIC|
```

Chaque variable est déclarée sous la forme :

<numéro de variable>**,**libellé en toutes lettres

et chaque modalité sous la forme :

<numéro de variable>**,**<numéro de modalité>**,** libellé en toutes lettres

Une fois que vous savez ça, il devient possible de remplir beaucoup plus vite le dictionnaire, tout particulièrement lorsque vous avez beaucoup de modalités à saisir. C'est notamment le cas lorsque vous voulez renseigner l'âge ou la PCS des répondants. Pour permettre une utilisation analytique des variables, il est bon de mettre en classe les réponses, afin de ne pas avoir une modalité différente pour chaque répondant. Par exemple, il est préférable d'avoir des classes d'âges plutôt que des âges exacts. Si vous avez des scrupules à ne pas garder l'information exacte, vous pouvez toujours l'ajouter parmi les observations rattachées à l'entretien, que vous pouvez remplir dans la page d'informations détaillées sur l'entretien.

Une méthode assez pratique pour construire les dictionnaires contenant beaucoup de catégories consiste à utiliser un tableur. Il suffit de placer chaque information dans une colonne (variable, modalité, libellée) et de concaténer le tout en bout de ligne. Dans l'exemple ci-dessous, on peut voir comment créer simplement le dictionnaire de la variable PCS au niveau 2 de l'Insee (les intitulés ont été recopiés directement en ligne):

Classeur1 - Microsoft Excel

NOMPROPRE =A13&","&B13&","&C13

	A	B	C	D
1	Var	Mod	Libellé	Concaténation
2	2	10	Agriculteurs exploitants	2,10,Agriculteurs exploitants
3	2	21	Artisans	2,21,Artisans
4	2	22	Commerçants et assimilés	2,22,Commerçants et assimilés
5	2	23	Chefs d'entreprise de 10 salar	2,23,Chefs d'entreprise de 10 salariés ou plus
6	2	31	Professions libérales et assir	2,31,Professions libérales et assimilés
7	2	32	Cadres de la fonction publique	2,32,Cadres de la fonction publique, professions intellectuelles et artistiques
8	2	36	Cadres d'entreprise	2,36,Cadres d'entreprise
9	2	41	Professions intermédiaires de	2,41,Professions intermédiaires de l'enseignement, de la santé, de la fonction publique et as
10	2	46	Professions intermédiaires adr	2,46,Professions intermédiaires administratives et commerciales des entreprises
11	2	47	Techniciens	2,47,Techniciens
12	2	48	Contremaîtres, agents de maît	2,48,Contremaîtres, agents de maîtrise
13	2	51	Employés de la fonction publi	=A13&","&B13&","&C13
14	2	54	Employés administratifs d'ent	2,54,Employés administratifs d'entreprise
15	2	55	Employés de commerce	2,55,Employés de commerce
16	2	56	Personnels des services direct	2,56,Personnels des services directs aux particuliers
17	2	61	Ouvriers qualifiés	2,61,Ouvriers qualifiés
18	2	66	Ouvriers non qualifiés	2,66,Ouvriers non qualifiés
19	2	69	Ouvriers agricoles	2,69,Ouvriers agricoles
20	2	71	Anciens agriculteurs exploitar	2,71,Anciens agriculteurs exploitants
21	2	72	Anciens artisans, commerçants,	2,72,Anciens artisans, commerçants, chefs d'entreprise
22	2	73	Anciens cadres et professions	2,73,Anciens cadres et professions intermédiaires
23	2	76	Anciens employés et ouvriers	2,76,Anciens employés et ouvriers
24	2	81	Chômeurs n'ayant jamais trava	2,81,Chômeurs n'ayant jamais travaillé
25	2	82	Inactifs divers (autres que re	2,82,Inactifs divers (autres que retraités)
26	2	10	Agriculteurs exploitants	2,10,Agriculteurs exploitants

Il suffit ensuite de recopier le contenu de la colonne D entre les balises |DIC| du fichier corpus pour que le dictionnaire soit rempli (il faut bien sûr penser à ajouter aussi la variable PCS entre les balises POS : ici « 2,PCS »).

Dans la mesure où vous avez jusqu'à 99 variables possibles, et que Sonal n'est pas forcément extrêmement souple en matière de recodage (pour l'instant), il est tout à fait possible de mettre plusieurs niveaux d'encodage pour une même variable. Par exemple, vous pouvez stocker l'âge en classe quinquennales et décennales (ci-contre).

DémoDico.xlsx - Micr

F43

	A	B	C	D	E
25		2	82	Inactifs divers (autres que re	2,82,Inactifs divers (autres que retra
26		2	10	Agriculteurs exploitants	2,10,Agriculteurs exploitants
27					
28		3	1	Moins de 15 ans	3,1,Moins de 15 ans
29		3	2	15 à 19 ans	3,2,15 à 19 ans
30		3	3	20 à 24 ans	3,3,20 à 24 ans
31		3	4	25 à 29 ans	3,4,25 à 29 ans
32		3	5	30 à 34 ans	3,5,30 à 34 ans
33		3	6	35 à 39 ans	3,6,35 à 39 ans
34		3	7	40 à 44 ans	3,7,40 à 44 ans
35		3	8	45 à 49 ans	3,8,45 à 49 ans
36		3	9	50 à 54 ans	3,9,50 à 54 ans
37		3	10	55 à 59 ans	3,10,55 à 59 ans
38		3	11	60 à 64 ans	3,11,60 à 64 ans
39		3	12	65 à 69 ans	3,12,65 à 69 ans
40		3	13	70 à 74 ans	3,13,70 à 74 ans
41		3	14	75 à 79 ans	3,14,75 à 79 ans
42		3	15	80 ans et plus	3,15,80 ans et plus
43					
44		4	1	Moins de 20 ans	4,1,Moins de 20 ans
45		4	2	20 à 29 ans	4,2,20 à 29 ans
46		4	3	30 à 39 ans	4,3,30 à 39 ans
47		4	4	40 à 49 ans	4,4,40 à 49 ans
48		4	5	50 à 59 ans	4,5,50 à 59 ans
49		4	6	60 à 69 ans	4,6,60 à 69 ans
50		4	7	70 à 79 ans	4,7,70 à 79 ans
51		4	8	plus de 80 ans	4,8,plus de 80 ans
52					
53					

Si vous ne souhaitez pas "fermer" la base de données en amont, utilisez la zone dévolue aux observations sur chaque entretien pour stocker les informations. Une

fois tous les entretiens réunis à nouveau, il sera possible de remplir la base données en une seule fois.

20.2 Distribuer le corpus modèle

Une fois le fichier corpus correctement construit, vous pouvez le distribuer aux différents participants de l'enquête. Pour ce faire, il suffit de leur adresser le fichier Crp que vous avez construit. C'est un simple fichier texte extrêmement léger que vous pouvez par exemple envoyer par mail.
Chacun pourra ensuite le recopier sur son ordinateur - idéalement dans le dossier qui contient leurs bandes - l'ouvrir et y charger ses propres entretiens pour les retranscrire et/ou les encoder.
En travaillant dans ce corpus, les différents intervenants vont progressivement enrichir les fichiers Rtr associés à leurs entretiens.

20.3 Centraliser les fichiers

Une fois qu'un chercheur a terminé de traiter un entretien (retranscrire/encoder thématiquement), il peut adresser le fichier Rtr à la personne en charge de centraliser les données.

- Comment réintégrer un fichier externe au corpus central?

Si l'équipe souhaite construire un corpus audio-textuel, chaque chercheur doit à la fois envoyer son fichier son (idéalement dans un format compressé) et le fichier Rtr associé à cette bande. La personne qui centralise les données colle ces deux fichiers dans le dossier du corpus collectif et ajoute le fichier son à l'intérieur du corpus. Comme Sonal ouvrira par défaut le fichier Rtr associé à la bande, l'entretien apparaîtra tel qu'il était dans le corpus du chercheur qui l'a encodé.

Si l'équipe ne souhaite pas utiliser la bande, il est également possible d'ajouter directement et uniquement le fichier Rtr au corpus. Celui-ci sera alors considéré comme un fichier texte et affiché comme tel. On perd certaines possibilités mais la manœuvre est plus légère.

21 Débuguer un corpus

Malgré tous les efforts faits pour stabiliser l'application et pour limiter les risques de perte de données, il est toujours possible qu'à un moment donné, vous ayez à débuguer votre corpus. Votre chance en l'occurrence est que le fonctionnement de Sonal est particulièrement simple. Tout repose sur des fichiers texte avec un système de balisage rudimentaire. Il est donc toujours possible de corriger manuellement ce qui dysfonctionne. Mais, pour ce faire, la première chose à faire est de se pencher un peu sur le fonctionnement du logiciel pour en comprendre l'organisation d'ensemble.

21.1 Diagnostiquer l'origine du problème

Pour faire simple, vous pouvez avoir deux types de problème : soit votre corpus contient une erreur et ne veut plus s'ouvrir (assez rare), soit il y a à l'intérieur de votre corpus un entretien qui dysfonctionne, c'est-à-dire un entretien où par exemple il y a des informations incohérentes etc. Les messages que Sonal vous envoient vous permettent généralement de faire la différence.

- Vérifier le fichier corpus

Si Sonal ne vous mentionne pas un entretien en particulier et qu'il se plante au démarrage d'un corpus, la première chose à faire est de vérifier l'intégrité du fichier corpus. Pour ce faire, il suffit de l'ouvrir avec le bloc note (clic droit -> ouvrir avec -> bloc note) et de vérifier qu'il est correctement structuré.
Y'a-t-il toutes les balises d'ouverture et de fermeture ? N'y-a-t-il aucun saut de ligne superflu ?
Il arrive par exemple que les lignes des thématiques contiennent des sauts de ligne intempestifs ou des lignes vides:

```
<|THEM|
 000,&H0, Pas de thématique
 001,16744576,Thématique 1
 004,16744448,Thématique
 Sur
 Plusieurs
 lignes
>|THEM|
```

En pareil cas, Sonal ne retrouve pas ce qu'il cherche sur chaque ligne (code, couleur, nom) et peut planter au moment du chargement des thématiques.
Au besoin, n'hésitez pas à faire des copies du fichier corpus (que vous placerez dans le même dossier que l'original) et à tenter des modifications.

Une bonne méthode pour s'assurer que le problème ne provient pas du corpus est de faire une copie du fichier crp, et d'y retirer tous les entretiens, c'est-à-dire de supprimer les lignes entre les balises **|ENTR|.**

Par exemple, votre fichier corpus contient :

```
<|ENTR|
 Entretien1.wav
 Entretien2.mp3
 Entretien3.wma
>|ENTR|
```

Il ne devra plus contenir que :

```
<|ENTR|
>|ENTR|
```

> *Ce faisant, vous ne supprimerez absolument pas vos données, vous vous contenterez de ne plus référencer les entretiens en question dans le corpus. Ceux-ci resteront disponibles pour une importation ultérieure.*

Après enregistrement, Si Sonal peut l'ouvrir, c'est bien que le problème provient d'un entretien en particulier, pas des paramètres du corpus.

Si jamais vous avez un doute sur l'entretien fautif. Vous pouvez repartir de votre corpus sans entretien et rajouter les entretiens (idéalement par un ajout collectif). Au moment du chargement des entretiens dans le corpus, vous pourrez voir au niveau de quel entretien se produit le problème.

21.2 Si un entretien dysfonctionne

Le problème le plus fréquent avec les corpus vient d'un entretien présentant des incohérences. En pareil cas, vous avez deux solutions : soit vous remplacez le fichier Rtr de votre entretien par une version antérieure (qui aura été sauvegardée soit manuellement, soit automatiquement), soit vous ouvrez le fichier Rtr (avec le bloc note) et vous le réparez manuellement (plus risqué…).

- Restaurer une version antérieure d'un entretien

Sonal crée un certain nombre de copies des fichiers Rtr lorsque vous travaillez. La première chose à faire est d'aller voir dans les copies faites automatiquement s'il n'y aurait pas une version stable du fichier Rtr que vous pourriez restaurer.

Pour cela, ouvrez le dossier « Restaur » situé dans le répertoire du projet.

Dans ce dossier, il y a au maximum 10 versions de chaque fichier Rtr. A chaque enregistrement, Sonal en crée un nouveau. **Le plus récent est celui dont l'index est le plus élevé**, en général le numéro 10, qui se présente sous la forme NomEntretien(10).Rtr.

Pour tenter une restauration, il vous faudra simplement recopier le fichier dont le nom se termine par « (10) » et le coller dans le répertoire du projet en lieu et place du Rtr effectivement utilisé dans le corpus.

> *Cela implique de d'abord renommer le Rtr de l'entretien défectueux situé dans le dossier projet. Passer par exemple de « NomEntretien.Rtr » à « NomEntretien(Ancien).Rtr » puis de coller le fichier Rtr issu du dossier Restaur à sa place, en lui donnant exactement le même nom, c'est-à-dire en retirant le « (10) » situé à la fin.*

Si, au rechargement du corpus, votre entretien bloque toujours. Essayez avec la version 9, puis la 8 et ainsi de suite jusqu'à retrouver une version stable. Evidemment, plus vous remonterez en arrière, plus vous risquez de perdre du travail.

Si vous possédez un dossier de sauvegarde récent, il vous suffira d'aller y recopier une version stable du Rtr de l'entretien fautif et de la coller à la place du Rtr situé dans le dossier du projet (pensez à renommer le Rtr fautif avant, on ne sait jamais…)

- Réparer manuellement un entretien

Ouvrir les fichiers Rtr avec le bloc note pourra vous permettre de solutionner un certain nombre de problèmes, notamment en lien avec les coordonnées des extraits. Reportez-vous aux explications détaillées sur le fonctionnement des fichiers Rtr pour vérifier la conformité des informations qu'il contient.

> *On peut par exemple corriger les extraits qui ont été ramenés par Sonal à deux secondes ou disjoints involontairement. Il suffit pour cela de recopier la balise de coordonnées de début de l'extrait suivant (par exemple « ::156,5 :: ») et de la coller en lieu et place des coordonnées de fin de l'extrait courant. Ainsi, la fin de l'extrait courant sera accolée au début du suivant.*

22 Traduire Sonal

Pour traduire Sonal, il faut choisir une langue dans le menu déroulant disponible au démarrage du logiciel :

Il vous faudra redémarrer le logiciel. Les principales fonctions seront traduites dans la langue voulue.

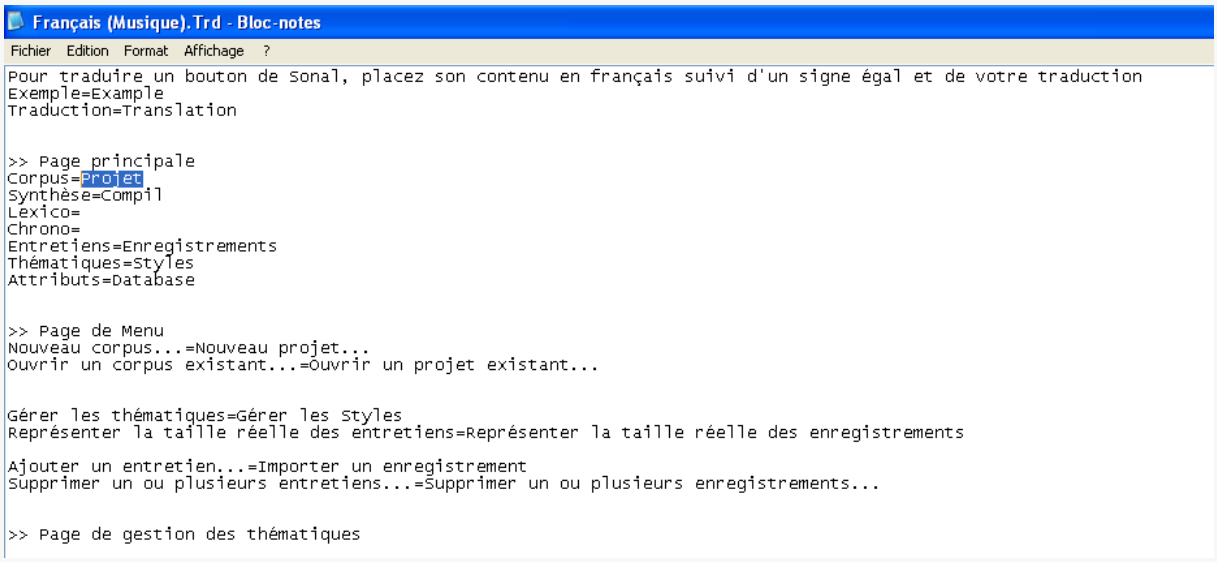
22.1 Sonal pour musiciens

Sonal peut être utilisé pour une grande variété de tâches en lien avec le son ou l'image. Pour répondre aux besoins de mon ami Grégoire Georgiou, bassiste de son état, une « traduction » de Sonal a été réalisée pour que les fonctions du logiciel soient plus adaptées au travail d'enregistrements musicaux et que les fonctions inutiles soient mises en retrait.
Les thématiques sont appelées styles, les entretiens des enregistrements et les synthèses des compils.

22.2 Personnaliser une traduction

Le système de traduction de Sonal est assez simple. Au chargement du logiciel, Sonal cherche les fichiers *.Trd qui sont contenus dans le dossier Sonal situé dans le répertoire de programmes de Windows (Program Files en général).

Ces fichiers sont de simples fichiers texte qui contiennent le texte de base du logiciel suivis de = et de leur traduction. Remplacez ce qui suit le = par votre traduction :



Enregistrez et rechargez Sonal. La traduction, si vous l'avez choisie, s'appliquera.

23 Informations

Les mises à jour sont chaudement recommandées. Le logiciel gagne en stabilité au fil du temps, il est dommage de s'en priver. **Dans tous les cas, les mises à jour ne peuvent pas nuire à vos données.**

Vous pouvez suivre le développement du logiciel en vous rendant sur son site ou sur le carnet de recherche qui lui est consacré sur hypothèses.org.

Si vous voulez suivre les mises à jour, vous pouvez vous abonner au flux RSS de Sonal.

Si vous voulez nous faire part de vos difficultés ou si vous avez des suggestions, vous pouvez nous écrire à cette adresse :

sonalteam@googlemail.com