

Leçon n°1 : Je découvre les dictionnaires de Prospéro

CE QUE VOUS ALLEZ APPRENDRE :

- *Constituer un projet et l'enregistrer*
- *Afficher les listes d'entités, d'épreuves, de qualités...*
- *Comprendre la structure des dictionnaires*

1. CONSTITUER UN PROJET

Travailler avec Prospéro suppose de commencer par lui donner un ensemble d'informations, enregistré sous un seul nom : c'est ce que l'on appelle un *projet*.

Ce fichier, dont l'extension est PRC, rassemble :

- la liste des textes composant le corpus (en fait, il faut préciser, les adresses des fichiers contenant les textes)
- l'adresse des dictionnaires (en première approche, considérons que les dictionnaires rassemble la liste des mots du corpus, de façon organisée puisque l'on distingue les noms, des verbes...)
- les adresses contenant le cadre d'analyse (c'est-à-dire les fichiers rassemblant les collections, les catégories et les êtres-fictifs).

Pour comprendre le fonctionnement de Prospéro, le plus simple est sans doute de créer un petit corpus de deux textes.

Avec votre traitement de texte habituel (ou avec WordPad, qui est livré avec Windows), vous allez taper les deux petits textes qui suivent. Attention, il faudra les enregistrer au format « texte » (l'extension sera .TXT). Ce format de fichier permet de ne garder que les caractères composants le texte et de ne pas polluer le fichier avec d'autres type d'informations enregistrées lorsque vous sauvegardez votre document « normalement » (par exemple, le type de police, les styles, la taille des marges...).

Attention, l'encodage du texte doit être en ISO 8859-1 (latin-1)

Pour enregistrer au format texte, il suffit de choisir cette option lorsque vous enregistrez votre document.



Voici les deux petits textes que je vous propose de taper (je vous invite à les enregistrer respectivement sous les noms de TEXTE1 et TEXTE2) dans un répertoire que vous dédierez à ces premiers exercices :

Texte 1

Ce texte est très heureux d'être le premier à figurer dans votre corpus. Cette antériorité lui procure un sentiment comparable à celui des pionniers débarquant sur une terre inconnue.

Puis le second :

Texte 2

Ce texte est un essai pour comprendre les dictionnaires de Prospéro. Il contient des mots classiques mais aussi des mots n'appartenant pas forcément aux dictionnaires de bases comme aphteuse, pilotis, stéroï de et sterbr ou encore des expressions comme « tennis de table ». On peut aussi ajouter un mot n'existant pas encore, comme par exemple, azertyu.

Apprenons maintenant à construire notre premier projet.

- Ouvrir Caliban A présent le fichier se nomme : prospero-I.exe
- Dans le menu **Fichier**, choisir **Gestion du projet**

Apparaît alors une boîte de dialogue comme celle-ci :



Commençons par préciser les textes composant notre corpus. Nous avons deux méthodes :



Comment faire pour ajouter un ou des textes à un corpus

1^{ère} méthode :

Cliquer sur le bouton **Ajouter un texte**

Il vous suffit alors de préciser l'endroit où le TEXTE1 se trouve et de cliquer sur **Ouvrir**. Il faut alors recommencer l'opération autant de fois que votre nombre de textes.

2^{ème} méthode :

Faire apparaître à l'écran simultanément l'explorateur Windows et la boîte de dialogue « gestion de projet » de Prospéro.

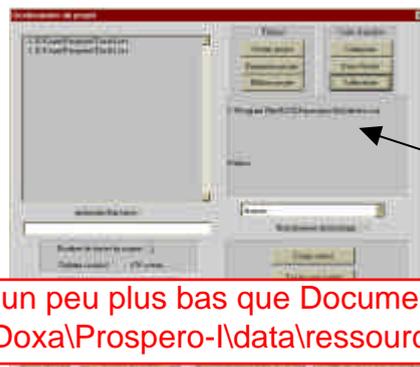
Avec l'explorateur, sélectionner les textes que vous souhaitez ajouter

Faire glisser les petites icônes correspondantes dans la fenêtre de gauche

A la fin de ces opérations, la liste des textes composant votre corpus fait apparaître celui (ou ceux) que vous venez d'ajouter.

Pour éviter de perdre ce travail, il est conseillé de l'enregistrer. En cliquant sur le bouton **Enregistrer sous** dans la fenêtre précédente, vous pouvez indiquer le nom de fichier de ce projet (l'extension, PRC, s'ajoute automatiquement) et surtout le répertoire dans lequel vous souhaitez le voir figurer. Mais votre définition du projet n'est pas terminée pour autant. Il vous faut à présent indiquer le *cadre d'analyse*, c'est-à-dire préciser les adresses des dictionnaires dans lesquels figurent les collections, catégories...

En appuyant sur le bouton **Catégories**, vous ouvrez une boîte de dialogue permettant de charger votre dictionnaire de catégories. Cette opération suppose que vous cliquez sur le bouton **Ouvrir**. Les dictionnaires de catégories se distinguent par l'extension CAT. Ouvrez par exemple le fichier ALERTES.CAT, puis fermez cette fenêtre en cliquant sur la petite croix, en haut à droite. La boîte de dialogue « Projet » fait apparaître l'adresse du fichier choisi.



Ici !

les fic/cat/col sont installés un peu plus bas que Documents\Doxa\Prospero sous C:\Users\XYZ\Documents\Doxa\Prospero-1\data\ressources\fr\concepts

Recommencez cette opération pour les Etres-Fictifs, puis pour les Collections (ces dictionnaires se distinguent par leurs extensions, respectivement, FIC et COL). Selon le même procédé, ouvrez donc les fichiers ALERTES.FIC et ALERTES.COL.

Enregistrez à nouveau votre projet en remplaçant le précédent.

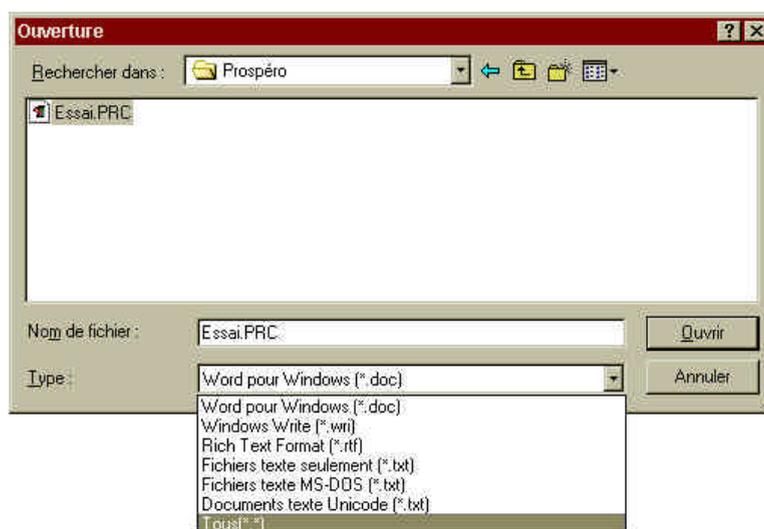
Vous venez de réaliser votre premier projet. Afin de bien comprendre la nature des opérations que vous venez de réaliser, je vous propose lire votre ESSAI.PRC.

Pour cela, il vous suffit de procéder ainsi :



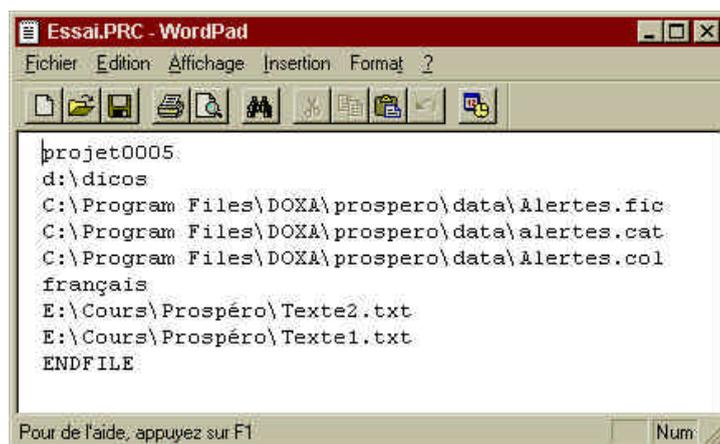
Comment faire pour éditer le contenu d'un projet

- Ouvrir votre traitement de texte (ou un éditeur comme WordPad ou le Bloc Notes de Windows)
- Lancer la commande **Fichier-Ouvrir**
- Ouvrir les répertoires appropriés pour identifier votre fichier. Vraisemblablement, vous ne trouverez pas car les traitements de texte n'affiche par défaut que les documents portant une extension particulière (par exemple DOC pour Word). Pour faire apparaître la liste des projets Prospéro, il vous suffit de taper *.PRC dans la zone « Nom de fichier » et de cliquer sur Ouvrir. Vous pouvez aussi faire apparaître l'ensemble des fichiers contenus dans le répertoire en choisissant « tous les fichiers » dans la partie « Type » :



- Une fois votre fichier identifié, cliquer sur Ouvrir.

Apparaît alors le fichier :



La première ligne contient une information qu'il ne faut surtout pas modifier.

La seconde fait apparaître le nom d'un dictionnaire important (nous y reviendrons)

Puis figurent les trois adresses des fichiers que vous avez indiquées en ouvrant successivement les dictionnaires des êtres-fictifs, des catégories et des collections.

Prospero a identifié que la langue des textes de notre corpus était le Français.

Viennent alors la liste des textes constituant notre corpus (ici nous en avons deux)

La commande ENDFILE clôt le document.

La lecture de ce document est précieuse, et ce, à plusieurs égards.

Tout d'abord, l'on comprend ainsi la nature d'un projet Prospero. Il ne contient essentiellement que des adresses (qu'il s'agisse des textes à analyser ou des dictionnaires permettant d'opérer).

Du coup, on peut imaginer taper directement sur votre éditeur ces renseignements, en prenant bien soin d'ajouter les commandes comme « Projet0005 », ou « ENDFILE ».

Mais surtout, l'édition du projet permet de modifier des adresses à moindre coût. Imaginons, que votre corpus contienne quelques 500 textes, stockés dans un répertoire sur un disque donné. En décidant de déplacer ce dossier ou les textes à un autre endroit de votre disque (voire sur un autre lecteur), il vous faudra indiquer à Prospero les nouvelles adresses. Si vous souhaitez éviter de recommencer l'ensemble des opérations que vous venez de réaliser, il vous suffit d'éditer le fichier dont l'extension est PRC. Avec votre traitement de texte, il vous est alors possible de **Remplacer** (commande se trouvant vraisemblablement dans le menu **Edition**) tous les *E:\Cours\Prospero* par *D:\Corpus\Essai*. Du coup, en pressant le bouton **Remplacer Tout**, vous indiquez en une seule fois les nouvelles adresses de vos textes.

2. TRAITEMENT, TYPAGE ET CONSTRUCTION D'EXPRESSIONS

Revenons à Prospéro. Il s'agit maintenant de lui demander de travailler, c'est-à-dire d'analyser votre corpus de texte. Il procède en deux temps.

A) Il va parcourir chaque texte et s'employer à noter le résultat de son travail dans un fichier portant le nom du texte, suivi de l'extension BDT. Tant que les dictionnaires ne changent pas, ce travail est réalisé une seule fois.

B) Il peut alors lire le résultat de son travail.

Pour lancer l'opération A, il suffit de se placer dans la zone « Retraitement automatique » et de cliquer sur le bouton **Corpus entier**. Celui intitulé **Textes non traités**, juste en dessous, est utilisé lorsque l'on vient d'ajouter quelques textes au corpus. Vous pouvez vérifier son travail en faisant apparaître avec l'explorateur Windows, le contenu du répertoire dans lequel figure TEXTE1.TXT et TEXTE2.TXT. Vous y trouverez désormais un TEXTE1.BDT et TEXTE2.BDT, résultats de cette opération. Ces fichiers sont éditables par votre traitement de texte, mais leur lecture reste difficile.

Pour bénéficier de cette phase d'analyse, il faut demander à Prospéro de lire ces fichiers (opération B). Le gros bouton, intitulé **Lecture des structures du corpus**, situé en bas de la boîte de dialogue, lance cette commande. Ici, elle ne dure qu'une fraction de seconde en raison de la taille de notre corpus (deux textes seulement). Elle sera beaucoup plus longue lorsque vous manipulerez des volumes de plusieurs centaines, voire de milliers, de textes.

Il est temps de commencer à examiner le travail réalisé par Prospéro. En lançant tour à tour les commande **Mode d'accès – Listes – Entités**, puis **Mode d'accès – Listes – Qualités**, **Mode d'accès – Listes – Marqueurs**,... vous obtenez le résultat des recensements effectués par Prospéro.

Nous reviendrons sur ces fenêtres. Comprendre comment fonctionne un logiciel, c'est s'interroger sur les ressources qu'il utilise. Certes, l'on voit bien l'intérêt de distinguer les entités, des qualités, des marqueurs... car sans cette distinction, il nous serait impossible de repérer comment les auteurs des textes introduisent de la variation pour qualifier, par exemple, un personnage. Mais comment Prospéro opère-t-il ? On pourrait imaginer qu'il dispose de dictionnaires contenant tous les noms, tous les adjectifs... Ce serait très lourd. En fait, les auteurs du logiciel ont fixé un ensemble de régularités se trouvant dans un fichier, nommé DICTIO.CFG. Vous pouvez l'éditer (il se trouve dans le répertoire C:\Program Files\DOXA\prospero) avec votre traitement de texte et repérer quelques unes de ces règles,

liées, on s'en doute, à la langue (quelques indications permettent de l'identifier). Cela ne signifie pas que vous pouvez constituer des corpus dans toutes les langues. Le travail nécessaire pour appréhender avec finesse une langue ne se fait pas en quelques mois. Mais lorsque l'on récupère de nombreux documents sur l'Internet, il faut bien identifier les langues des différents textes.

Il nous affranchit, nous l'avons compris, de reconnaître la majorité des mots du corpus. Cependant, un travail sur ces régularités se saurait suffire. D'une part, il est des mots n'obéissant pas aux règles. La terminaison de l'entité « nez », par exemple, suggère qu'il s'agit d'un verbe. D'autre part, la langue contient des expressions comme « le Tour de France », « bruit de fond », « tennis de table »... que l'on ne saurait réduire à la juxtaposition de mots. Il faut donc *créer (ou construire) l'expression* et la *typer*.



Définitions

Créer (ou construire) une expression, c'est indiquer à Prospéro d'isoler une séquence de mots (ou de signes) et lui indiquer qu'il s'agit d'un seul bloc.

Typer un mot (ou une expression), c'est indiquer s'il s'agit d'une entité, d'une qualité, d'un marqueur.

Commençons par typer les mots « indéfinis », c'est-à-dire dont le type n'a pas pu être reconnu grâce à la terminaison. En lançant la commande **Mode d'accès – Listes – Indéfinis**, on obtient la fenêtre suivante.



Le mot « Texte » n'a pas été identifié comme une entité et est absent de ce dictionnaire. En cliquant sur le bouton **Typier**, on ouvre une fenêtre permettant d'indiquer sa nature :

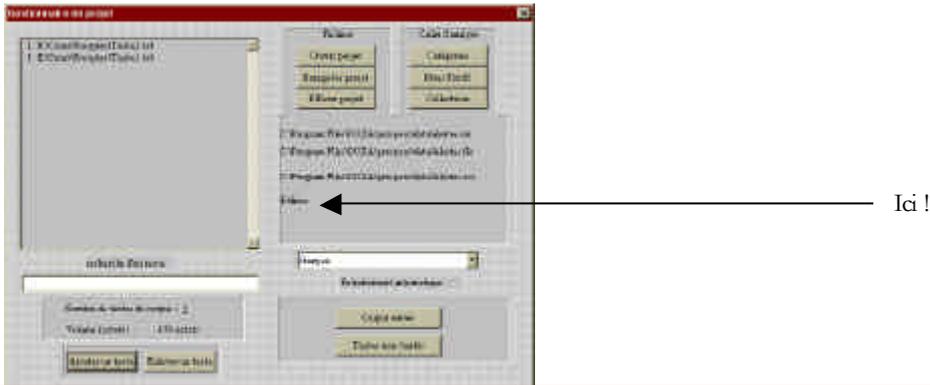


En cliquant sur OK, « Texte » perd son statut « d'indéfini » pour prendre celui d'entité. Pour s'en convaincre, fermez la fenêtre des indéfinis (grâce à la petite croix en haut à droite), et ouvrez la liste des entités. Vous pouvez constater que « Texte » figure dans cette fenêtre.

Chez certains utilisateurs, le typage provoque une sortie intempestive du logiciel tant que les dictionnaires ne sont pas définis dans le projet. Reportez-vous aux instructions de la page suivante pour corriger ce problème.

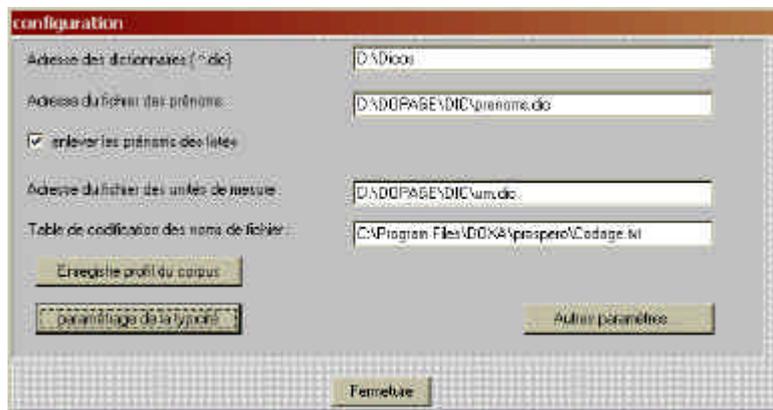
En la parcourant, vous avez sans doute noté que le mot « procure » a été considéré comme une entité. Ce n'est pas faux mais ici, il s'agit d'une forme conjuguée du verbe « procurer ». Il faut donc le « ReTypier », néologisme qui signifie que l'on souhaite corriger le type proposé automatiquement par Prospéro. Dans la liste des entités, sélectionnez le mot « procure ». Le bouton **ReTypage** vous permet de faire réapparaître la fenêtre ci-dessus, et donc d'indiquer qu'il s'agit d'une épreuve.

Examinons les conséquences de ces opérations. En typant un mot, on l'enregistre dans un dictionnaire correspondant à son statut. Il existe des fichiers FR_ETRE.DIC, FR_EPREU.DIC, FR_MARQU.DIC, FR_QUALI.DIC... qui contiennent respectivement la liste des entités, des épreuves, des marqueurs, des qualités,... typés « à la main » (c'est-à-dire sans prendre en compte la terminaison du mot). Typier une entité (respectivement une épreuve, un marqueur...) c'est faire figurer ce mot dans ce document. Ces fichiers (dont l'extension DIC indique qu'il s'agit de dictionnaires) doivent figurer dans le même répertoire, qui lui-même doit apparaître dans la fenêtre « Gestion de projet ».



les dictionnaires sont sous :
 C:\Users\XYZ\Documents\Doxa\Prospero-1\data\ressources\fr\dictionnaires

Pour modifier cette adresse, il faut lancer la commande **Fichier – Configuration du projet** et indiquer le chemin sur la première zone de la boîte de dialogue (adresse des dictionnaires).



Nous reviendrons plus tard sur les autres options de cette fenêtre.

Pour bien comprendre l'opération de typage, il faut sans doute ouvrir ces dictionnaires. La manipulation suivante devrait vous permettre de repérer le fonctionnement de Prospéro.

- Faire apparaître la liste des indéfinis (souvenez-vous... **Mode d'accès – Listes – Indéfinis**)
- Repérer un mot figurant dans cette fenêtre. Par exemple le mot « pilotis », n'est pas typé dans mon dictionnaire. Dans tous les cas, votre néologisme « azertyu » est vraisemblablement indéfini.
- Ouvrir avec votre traitement de texte (ou votre éditeur), le fichier FR_ETRE.FIC qui contient, rappelons-le, l'ensemble des entités qui ont été typées sans prendre en compte la terminaison.
- Vérifier que « pilotis » n'y figure pas (vous pouvez utiliser la fonction **Recherche** de votre traitement de texte ou de votre éditeur).
- Sous Prospéro, taper « pilotis » comme une entité.
- Vérifier que ce mot figure désormais dans la liste des entités (pour cela fermer la fenêtre puis lancer la commande **Mode d'accès – Listes – Entités**).
- Vérifier que ce mot figure maintenant dans le fichier FR_ETRE.DIC (il a sans doute été rajouté à la fin de la liste).

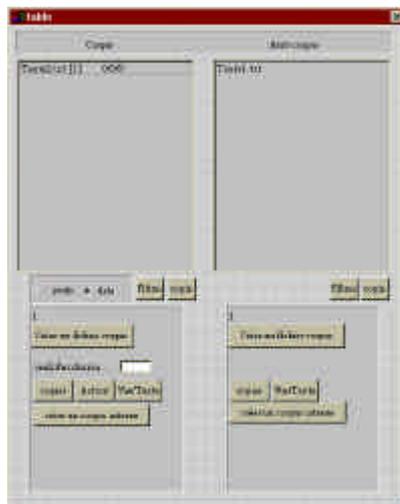
Apprenons à présent à construire une expression.

- Faites apparaître la liste des entités dans laquelle apparaît le mot table (souvenez-vous... dès qu'il s'agit d'une liste, il faut aller dans le menu **Mode d'accès – Listes - ...**

- Cliquer sur le mot « table »

- Appuyer sur le bouton **Enoncés**

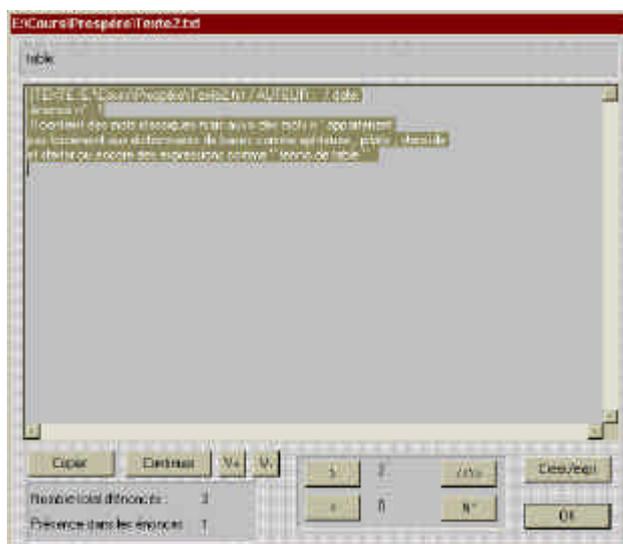
Une nouvelle fenêtre se présente :



On peut repérer deux zones. Sur la première (nommée « corpus »), on retrouve l'ensemble des textes dans lesquels figure le mot « table ». La seconde, devenant donc « l'anti-corpus » dresse la liste de ceux dans lesquels le mot est absent. Comme sur les autres boîtes de dialogue que vous avez repérées, oublions pour l'instant la majorité des boutons.

Cliquez sur TEXTE2.TXT, qui contient donc notre mot table, puis appuyez sur le bouton

Action. le bouton [action] a été renommé en [détail]



L'on repère l'énoncé dans lequel figure le mot « table ». Un énoncé, pour Prospéro, correspond en gros à une phrase. En fait, il s'agit de l'ensemble des mots figurant entre deux points (ou autres ponctuations de ce type comme le point d'interrogation, d'exclamation...).

L'opération que vous venez de réaliser est très importante car très courante. En effet, lors de votre travail avec Prospéro, à chaque instant, vous avez la possibilité de vous replonger dans le texte pour vérifier l'usage d'un mot, d'une expression, pour repérer les conditions d'émergence d'un personnage ou d'un sigle, pour valider une hypothèse... Nous souhaitons créer l'expression « tennis de table ». Finalisons cette opération.



Comment faire pour construire une expression

- Après avoir fait apparaître l'énoncé, sélectionner les mots composant l'expression.
- Cliquer sur le bouton **Créer-expr**
- Indiquer le type (ici, il s'agit d'une entité)
- Valider en appuyant sur le bouton OK
- Fermer les différentes boîtes de dialogue

Pour vous convaincre de l'opération, il suffit de vérifier que l'expression figure dans le fichier FR_ETRE.DIC. Ouvrir ce fichier avec votre traitement de texte ou avec votre éditeur. A la fin du document, vous devez voir apparaître tennis de table.

Cela signifie que Prospéro considèrera désormais que ces trois mots (« tennis », « de », « table »), lorsqu'ils se suivent, ne doivent plus être traités séparément mais ensemble, et comme une entité.

Dans la liste des entités, c'est tennis de table qui apparaîtra comme une seule entité. Cela ne marche pas ? C'est normal ! Car si « tennis de table » apparaît bien dans le dictionnaire, Prospéro n'a pas encore eu l'occasion de retraiter l'ensemble des textes pour identifier ce bloc de trois mots.

Retournez dans la fenêtre **Gestion du projet** accessible par le menu **Fichier**. Relancer le traitement automatique (**Corpus entier**). Puis cliquer sur **lecture des structures du corpus**. Si vous retournez à la liste des entités, vous verrez apparaître « tennis de table »... Si cela ne marche

toujours pas, il faut probablement effacer de la mémoire les anciennes informations. Pour cela, dans la boîte de dialogue **Fichier – Gestion du projet**, cliquez sur **Effacer** (n’avez pas peur, ce bouton revient simplement à fermer le projet). Ouvrez-le à nouveau, lancer le traitement automatique puis la **lecture des structures du corpus**. Si vous retournez à présent à la liste des entités, vous obtenez bien la preuve de la création de l’expression.



 **Récapitulons...**

Vous connaissez désormais plusieurs dictionnaires. Que contiennent-ils ? Cette petite fiche devrait vous permettre de mémoriser les éléments importants. Je commence à la compléter... finissez le travail (au besoin, ouvrez les documents pour confirmer vos doutes...)

ALERTE.FIC : Contient la liste des êtres fictifs (cette notion sera précisée lors de la leçon suivante). Son adresse est indiquée dans la fenêtre **Fichier – Gestion du projet**. Pour la changer, il suffit de cliquer sur le bouton approprié et d’ouvrir le fichier voulu.

ALERTE.COL : Contient la liste des collections (notion précisée lors de la leçon suivante). Même remarque que pour les êtres-fictifs concernant son adresse.

ALERTE.CAT : Contient la liste des catégories (notion précisée lors de la leçon suivante). Même remarque que pour les êtres-fictifs concernant son adresse.

DICTION.CFG :

.....

FR_ETRE.DIC :

Son adresse est indiquée dans la fenêtre

Pour la changer,

FR_EPREU.DIC :

(Même remarque pour l'adresse)

FR_QUALI.DIC

(Même remarque pour l'adresse)

FR_MARQU

(Même remarque pour l'adresse)

FR_OUTL

(C'est la même logique, ouvrez ce fichier pour compléter les lignes ci-dessus)

FR_NOMB

Figurent également des fichiers que l'on n'a pas encore étudiés :

UM.DIC contient l'ensemble des unités de mesure (voir leçon n° 3). Son adresse est indiquée dans la fenêtre accessible par la commande **Fichier – Configuration du projet**

PRENOM.DIC contient la liste de prénoms. Ce fichier, très important, nous aidera à identifier les personnes (cf. Leçon n° 3). Son adresse est également mentionnée dans la fenêtre **Fichier – Configuration du projet**

Il ne reste plus qu'un seul autre dictionnaire important : il s'agit d'un fichier permettant de préciser automatiquement certaines références des textes. Nous l'étudierons à la leçon n° 4.

Avant de passer à la suite, je vous encourage à réaliser quelques exercices.

3. EXERCICES

Exercice 1 :

Typier tous les indéfinis de votre projet.

Editer les dictionnaires correspondant pour vérifier vos opérations.

Exercice 2 :

Créer un nouveau répertoire sur votre disque dur, puis déplacer les fichiers TEXTE1.TXT, TEXTE2.TXT, TEXTE1.BDT, TEXTE2.BDT ainsi que ESSAI.PRC dans ce dossier. Modifier ce dernier fichier en lui indiquant les nouvelles adresses.

Exercice 3 :

Créer un répertoire « Dictionnaire » sur l'un de vos disques durs. Copier tous les fichiers nécessaires à Prospéro (dont l'extension est DIC) dans ce nouveau répertoire.

Faire les modifications nécessaires sur ESSAI.PRC.

Déplacer les dictionnaires de concepts (catégories, collections et êtres-fictifs). Faire les modifications nécessaires pour pouvoir continuer à travailler sur ESSAI.PRC.

Leçon n°2 : Je navigue dans le corpus

CE QUE VOUS ALLEZ APPRENDRE :

- *Automatiser la saisie des références externes*
- *Les commandes de base pour naviguer dans le corpus*
- *La gestion des concepts (être-fictif, collection et catégorie)*
- *La démarche du chercheur prospérien*

En remplacement du corpus Concorde, vous trouverez un corpus constitué de quelques textes de candidats aux élections présidentielles 2002 et 2007. Vous pouvez le télécharger à l'adresse :

http://prosperologie.org/dl.php?filename=premier_tour_2002-2007.zip

(nécessite d'être connecté et identifié sur le site, et donc de s'inscrire si nécessaire)

A dézipper ensuite dans C:\corpus\

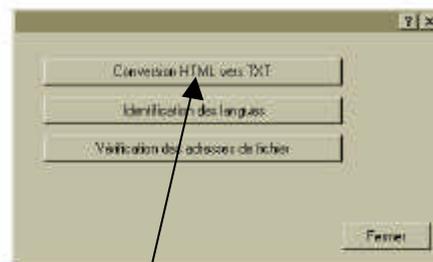
1. LES REFERENCES EXTERNES

Pour cette leçon, comme les suivantes, je vous propose de travailler sur le corpus Concorde. Vous disposez de l'ensemble des textes disponibles à ce jour. Lors de la leçon n°1, vous avez appris à construire un corpus. Belle occasion de reprendre vos notes pour réaliser cette opération : faire glisser l'ensemble des textes, indiquer les adresses des dictionnaires en ouvrant, tour à tour, les fichiers contenant les êtres-fictifs, les collections, les catégories,... Pensez à enregistrer le projet sous le nom de Concorde.PRC (l'extension s'ajoute automatiquement).

Il se peut que les textes mis à votre disposition soient au format HTML, format souvent utilisé lorsque l'on récupère des informations sur l'Internet. Il existe une procédure permettant de les convertir au format texte (TXT). Pour la lancer, il suffit de cliquer sur le bouton **Autres options** situé en bas à gauche de la boîte de dialogue **Fichier – Gestion du projet**. Une fenêtre s'ouvre. Il suffit alors de cliquer sur le bouton



Ici...



puis là !

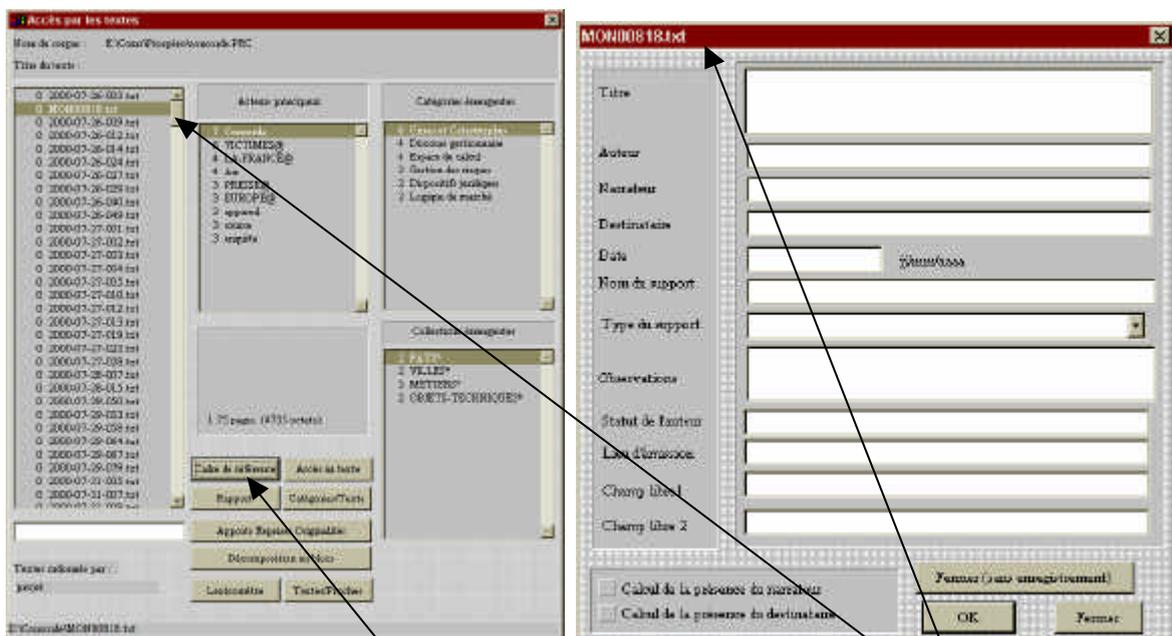
Vous pouvez alors demander le traitement automatique du corpus (qui revient, rappelons-le à créer les fichiers BDT), puis la lecture de cette opération. Vous pouvez alors faire apparaître les listes d'entités, de qualités,...

Nous allons apprendre dans ce chapitre à naviguer dans le corpus. Naviguer signifie, pour l'instant, recenser les termes les fréquents, repérer les expressions particulières, les collections et les catégories sur lesquelles s'appuient les auteurs... Mais il serait intéressant de savoir précisément quels sont les auteurs mobilisant essentiellement telle formule rhétorique, tel marqueur ou, d'une façon plus générale, telle ressource. Cette opération ne peut aboutir si l'on n'indique pas à Prospéro les auteurs des différents textes. De même, sans les dates des textes, il sera impossible de recenser non seulement les appuis historiques du discours, les ouvertures sur l'avenir et les modalités de la gestion de l'incertitude, mais aussi de repérer un événement marquant, un élément précurseur, les silences du dossier (c'est-à-dire des périodes pendant lesquelles nous ne disposons d'aucun texte).

Bref, nous l'aurons compris, pour chaque document du corpus, il nous faut renseigner une petite fiche indiquant les références, dites «externes» car ces informations ne sont pas directement contenues dans le texte lui-même. Le paragraphe précédent souligne l'importance de l'auteur et de la date, mais l'on pourrait, selon les cas, accorder beaucoup d'attention au type de support, au

statut de l'auteur, ou à des informations pertinentes selon le dossier étudié et la problématique du chercheur.

Les auteurs de Prospéro ont donc créé un système de base de données permettant de rassembler l'ensemble de ces informations. Pour faire apparaître les rubriques disponibles (dans le vocabulaire des bases de données, on parle de champs), il vous suffit de lancer la commande **Mode d'accès – Textes**. Oublions pour l'instant l'ensemble des informations accessibles par cette boîte de dialogue. Choisissez un texte sur la colonne de gauche et cliquez sur le bouton **Cadre de référence**. Vous obtenez alors le formulaire (on parle aussi de masque de saisie) permettant de renseigner les différentes rubriques pour le texte sélectionné.

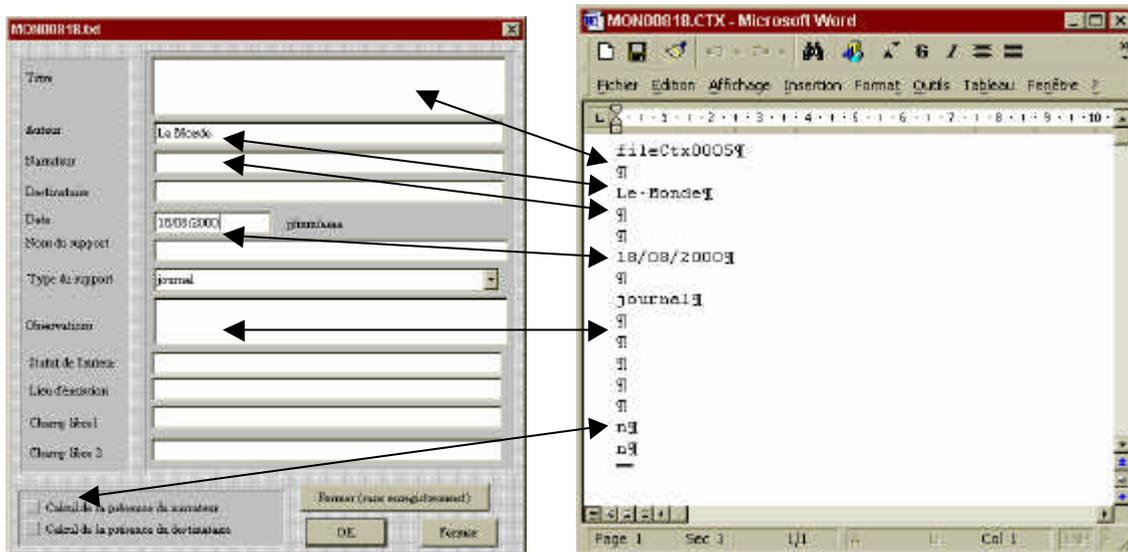


En cliquant ici, vous accédez au formulaire concernant le texte choisi

Ainsi, les champs de la boîte de dialogue de droite (Auteur, Destinataire,...) peuvent jouer le rôle de variables discriminantes (comme l'on dit dans les sociologies quantitatives) c'est-à-dire qui permettent de distinguer les résultats selon l'auteur, son statut,... Mais nous verrons que ces informations peuvent être utilisées pour avoir une idée du texte sans avoir à l'ouvrir, et surtout pour analyser l'épaisseur temporelle du dossier (champ date).

Tous ces projets nécessitent de renseigner l'ensemble de ces informations pour chacun des textes. Elles s'inscrivent dans un fichier dont l'extension est CTX. Il y aura donc autant de CTX que de

de la présence du narrateur, calcul de la présence du destinataire), les deux dernières lignes du CTX contiennent la valeur n (pour non).



Ainsi, apparaît la correspondance entre les informations saisies dans le formulaire et le fichier CTX. Cela signifie notamment que si vous savez programmer, vous pouvez générer directement vos CTX en utilisant un langage adapté au traitement de chaîne de caractères (par exemple le PERL). Comprendre la structure des fichiers CTX permet également de saisir ces informations sur un ordinateur ne possédant pas Prospéro. Il vous suffit pour cela de créer autant de fichier CTX que de fichiers TXT en prenant soin de les nommer de la même façon et de respecter l'ordre des champs (un retour chariot par rubrique). La saisie «manuelle» de toutes les informations pour chacun de nos 474 textes est cependant fastidieuse, mais parfois inévitable. Deux outils existent pour vous affranchir d'une partie de ce travail digne d'un cistercien.

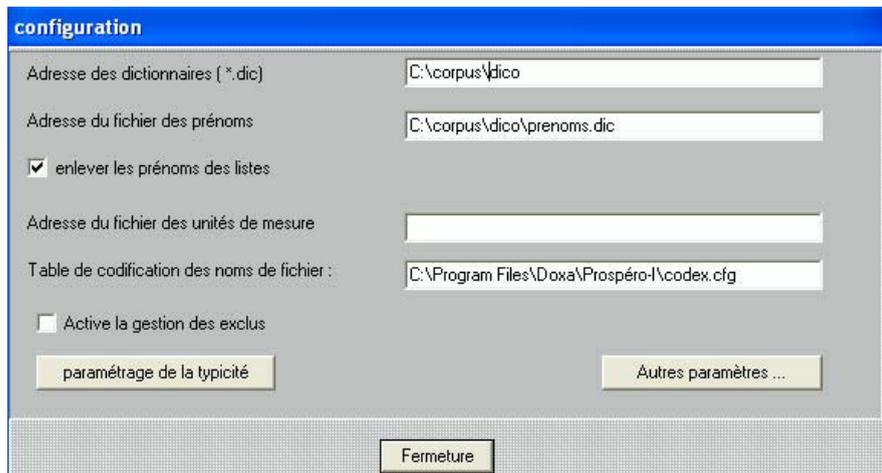
A) Le premier moyen de coder automatiquement les références externes revient à utiliser le nom du fichier.

Il faut procéder en trois temps. D'une part, il convient de donner un nom aux fichiers des noms «intelligents». La convention retenue invite à fixer quelques lettres indiquant une abréviation, la date (selon un format un peu particulier) et un texte libre. Par exemple, si votre fichier s'appelle AFP00714TOTO, Prospéro aura les moyens d'identifier que AFP désigne un sigle (nous verrons comment plus bas), que 00714 renvoie à une date et que TOTO est donc un nom libre...

Commençons par la date. Son format est AAMJJ, ce qui signifie qu'il faut commencer par donner 2 caractères pour l'année (01 pour 2001, 98 pour 1998), un caractère pour le mois et deux pour le

jour (01, 02, ... 31). Ce qui surprend le non-informaticien, c'est le mois. Coder sur un seul caractère le mois paraît surprenant même pour un enfant, dès que l'année est divisée en 12 mois. C'est par économie de place sur les vieux systèmes (sur un PC, les noms de fichiers étaient limités à 8 caractères avant l'arrivée de Windows 95), que cette convention avait été mise en place. L'on procède donc ainsi : de 1 à 9 pour les mois de janvier à septembre, puis A pour octobre, B pour novembre, C pour décembre.

Pour l'abréviation, il faut indiquer à quoi elle correspond dans un fichier au format TXT. Vous en avez sans doute une version de ce document dans le répertoire C:\Program Files\DOXA\prospero. Il s'appelle **codex.cfg** mais son nom est libre du moment qu'il est au format TXT. Il doit être référencé dans la fenêtre accessible par la commande **Fichier – Configuration du projet** : il suffit pour cela d'inscrire son nom et son adresse dans la zone « Table de codification des noms de fichier ».



Le fichier est lui-même composé **ainsi**

```

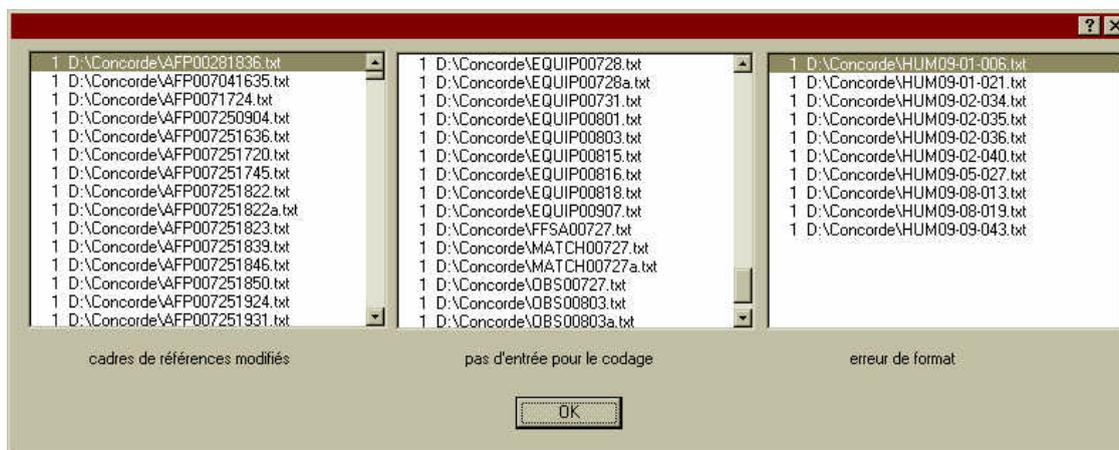
codex.cfg - Bloc-notes
Fichier Edition Format Affichage ?
#####
ABREV:AAAA
AUTEUR:Association des Authentiques Amateurs d'Andouillettes Ardéchoises
SUPPORT:Bulletin de l'AAAAA
TYPE-SUPPORT:Revue
STATUT-AUTEUR:Gourmet
LIEU-EMISSION:Ardèche
CHAMP-1:Attention aux graisses
CHAMP-2:Alimentation à risques sanitaires
OBSERVATION:Références externes de démonstration
#####
ABREV:Libé
AUTEUR:Libération
SUPPORT:Libération
TYPE-SUPPORT:Presse Quotidienne Nationale
STATUT-AUTEUR:Journaliste
LIEU-EMISSION:Paris
#####
ABREV:Lib
AUTEUR:Libération
SUPPORT:Libération
TYPE-SUPPORT:Presse Quotidienne Nationale
STATUT-AUTEUR:Journaliste
LIEU-EMISSION:Paris
#####
ABREV:60M
SUPPORT:60 Millions de consommateurs
TYPE-SUPPORT:News et Magazines
STATUT-AUTEUR: Presse consommériste
#####
ABREV:AFP
AUTEUR:AFP
SUPPORT:AFP
TYPE-SUPPORT:Agence de presse
STATUT-AUTEUR:Journaliste

```

Le plus simple pour ajouter une nouvelle source à coder automatiquement est de copier-coller une référence existante et de l'adapter à votre usage

Une fois que vos noms de fichiers sont homogènes, que vous disposez d'une table de codification comme celle-ci (déclarée dans **Fichier – Configuration du projet**), il vous suffit de lancer la commande **Codage des références externes** dans la fenêtre **Fichier – Gestion du projet**. Pour cet exercice vous pouvez vous contenter d'utiliser le **codex.cfg** disponible dans le répertoire C:\Program Files\DOXA\prospero.

Un message vous annonce que cette opération revient à détruire les anciens CTX qui existeraient préalablement. Puis s'affiche une fenêtre présentant le compte-rendu de l'opération. Elle se compose de trois espaces :



La colonne de gauche présente l'ensemble des textes pour lesquels le codage s'est effectué sans problème. Celle du milieu rassemble la liste des textes pour lesquels votre table de codification (ici, **codex.cfg**), ne contient pas l'abréviation utilisée. Par exemple dans cet exemple, mon fichier **codex.cfg** ne comportait pas de ligne pour OBS (Le Nouvel Observateur). Du coup, Prospéro ne peut pas coder ces textes. Enfin, dans la colonne de droite, nous obtenons une seconde liste d'erreurs. Le logiciel a certes identifié l'abréviation des textes mentionnés mais leur format n'est pas celui fixé. Ici, mon **codex.cfg** comporte bien une ligne pour HUM (L'Humanité) mais la date n'est pas sous la forme définie plus haut.

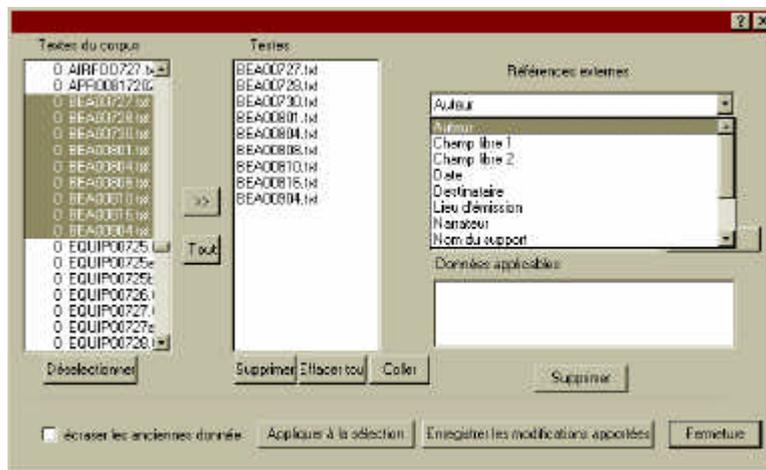
Il est possible que vous ne trouviez aucune erreur. Cela signifie que votre version du corpus et votre fichier **codex.cfg** ont été corrigés. Si tel n'est pas le cas, il vous appartient de le faire.

B) Un autre outil de codage des références externes.

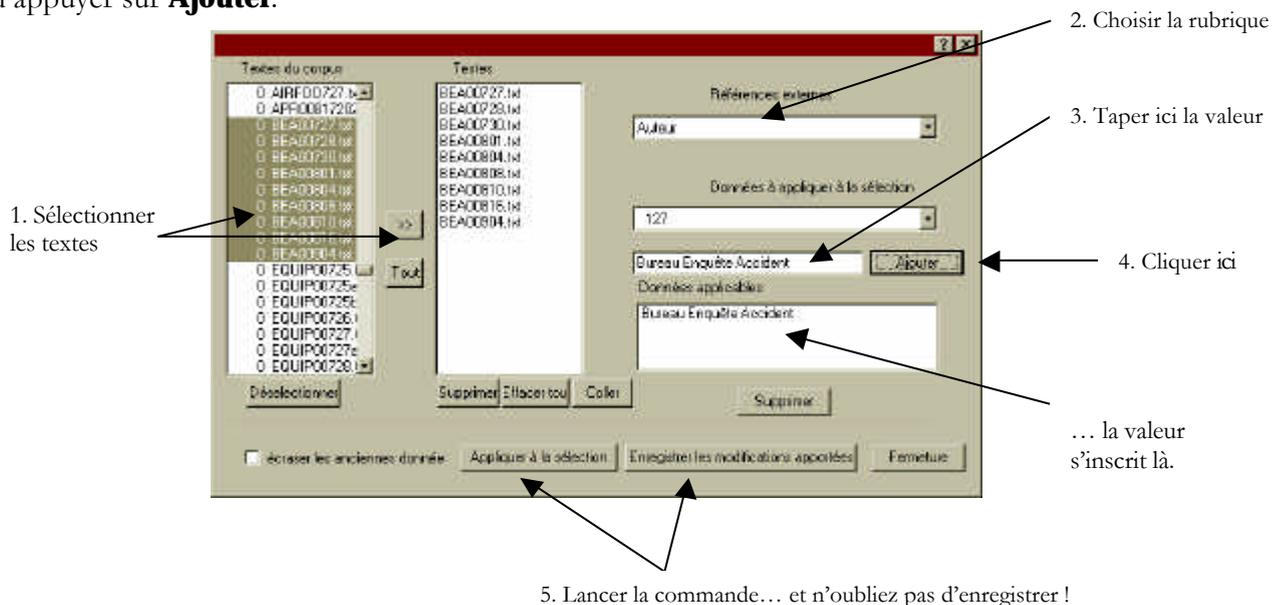
Il vous est aussi possible d'utiliser la commande **Outils – Recodage des références externes**.

Cette fonctionnalité peut s'employer pour compléter d'autres champs (oublié, par exemple, lors de la précédente opération). Une fois la commande lancée, il faut travailler en plusieurs temps. Tout d'abord, il convient de sélectionner les textes pour lesquels on souhaite ajouter une même information dans les références externes. On pourrait par exemple décider de compléter le champ « Auteur » avec la mention « Bureau Enquête Accident » pour tous les textes dont le nom commence par BEA.

En sélectionnant ces textes dans la première colonne de la fenêtre, on parvient à les déplacer dans la seconde en cliquant sur le bouton .



On décide alors de travailler sur la rubrique « Auteur » en choisissant le champ correspondant dans la zone «Références Externes». On tape la valeur que doit prendre la variable avant d'appuyer sur **Ajouter**.



Il ne vous reste plus qu'à vérifier les résultats de votre travail, soit lançant la commande **Cadre de référence** dans fenêtre **Mode d'accès – Textes**, soit en regardant directement les CTX avec votre traitement de texte.

Enfin, il est un champ facile à remplir : le titre. En effet, une solution simple est sans doute de prendre le début de chaque texte pour le faire figurer dans cette rubrique. Si elle vous convient, il vous suffit d'appuyer sur le bouton **Attribution automatique des titres** dans la fenêtre **Fichier – Gestion de projet**



Quelques remarques sur l'usage des références externes

Le néophyte s'interrogera peut-être : que faut-il mettre dans « observations » ? Faut-il tout remplir ?

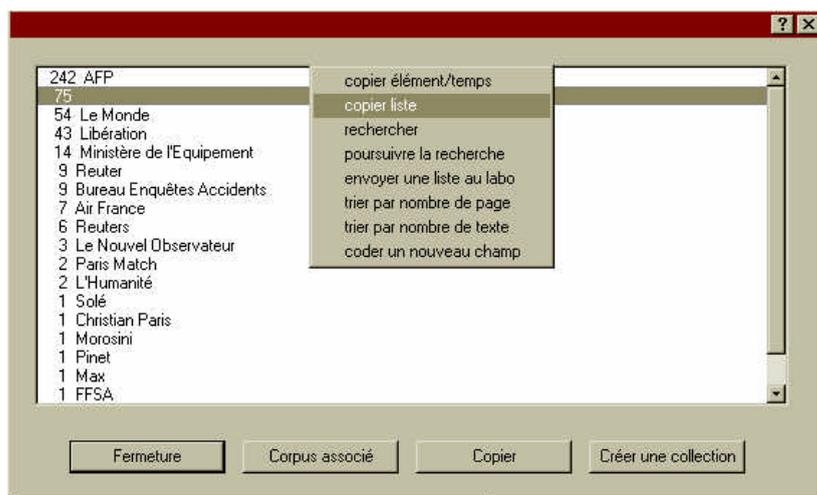
Il faut considérer les références externes comme des appuis que le chercheur pourra mobiliser pour faire des expériences. Seul le remplissage de la date est, de mon point de vue, incontournable. Compléter l'auteur et le type de support constitue un atout supplémentaire. Les autres champs peuvent être utiles en fonction des préoccupations du chercheur. Rien n'empêche d'employer la rubrique « observations » (ou l'un des champs libres) pour désigner le sexe, la CSP ou le parti politique de l'auteur si vous travaillez respectivement dans la tradition des *gender studies*, des sociologies de la distinction ou dans une perspective de sciences politiques.

La démarche pragmatique invite à compléter les champs en fonction des contraintes pesant sur les acteurs. Par exemple, vous pouvez indiquer dans l'une des rubriques mises à votre disposition, si oui ou non, tel auteur a eu connaissance d'un texte ou d'un fait.

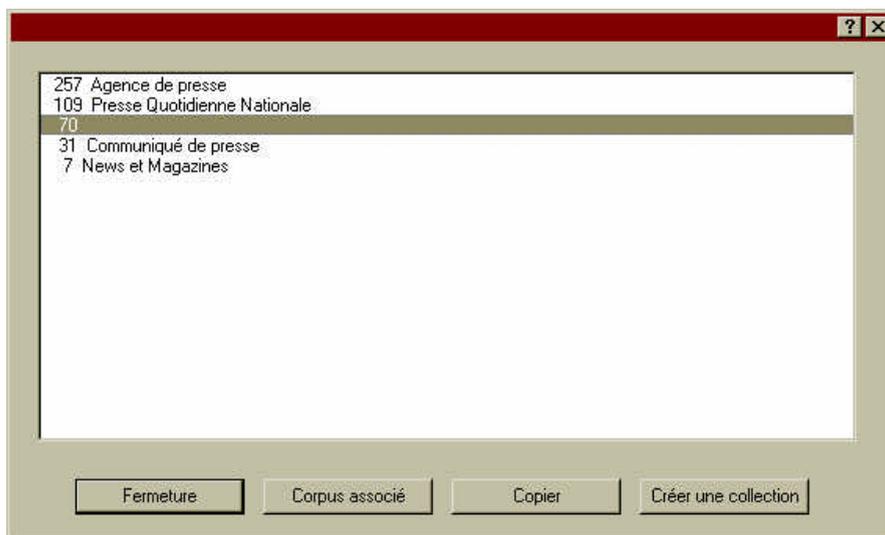
Dans tous les cas, l'on pourra évaluer le poids des variables discriminantes, en distinguant les résultats en fonction d'elles. La question est de savoir qui, du chercheur ou des acteurs, doit fournir les informations nécessaires pour compléter les variables... A ce titre, l'on se reportera avec intérêt à l'ouvrage de Francis Chateauraynaud : *Prospéro, une technologie littéraire pour les sciences humaines* (2002, A paraître)

Les références externes, vous l'aurez compris, servent donc surtout à discriminer les résultats en fonction des valeurs saisies dans les différents champs de cette base de données. Mais avant d'évaluer l'influence de telle ou telle variable, il est souvent utile de recenser leurs modalités. En clair, nous disposons de quelques outils permettant d'effectuer « le tri à plat » comme dirait un statisticien, c'est-à-dire de lister l'ensemble des valeurs prises pour étudier la distribution de la variable. Ces commandes sont accessibles par le menu **Mode d'accès – Appuis externes - ...**

Ainsi en lançant **Mode d'accès – Appuis externes – Auteurs**, vous obtenez dans une fenêtre, la liste des auteurs de votre corpus (en fait, ceux que vous avez bien sûr saisis) et le nombre de textes correspondant. Avec le bouton droit de la souris vous pouvez copier cette liste dans votre traitement de texte, mais aussi réaliser quelques opérations intéressantes. Le codage d'un nouveau champ, vous ouvre une boîte de dialogue semblable à celle qui vous a permis de renseigner automatiquement les références externes (**Outils – Recodage des références externes**)



Il en est de même pour les autres références externes que vous avez saisies. Par exemple, il est facile de faire apparaître la liste des types de support :



Comment avez-vous fait ?.....

Mais ces fenêtres permettent également d'automatiser la correction des erreurs dans les références externes. Admettons par exemple que, après une saisie manuelle des noms des auteurs, vous vous trouviez avec plusieurs auteurs différents (J. C. Gayssot, Gayssot, GAYSSOT...) pour désigner le Ministre des transports. Ces formes apparaissent dans la fenêtre **Mode d'accès – Appuis externes – Auteurs**. Il vous est possible de recoder rapidement, par exemple les deux dernières formes, en « J. C. Gayssot ». Comment ? Dans la boîte de dialogue précédente, choisissez GAYSSOT, puis cliquer sur **Corpus Associé**. Vous retrouvez alors une fenêtre bien connue. Appuyer sur le bouton **Copie** situé à droite du bouton **Filtre**. Il vous suffit alors de lancer la commande **Outils – Recodage des références externes** et de cliquer sur le bouton **Coller** (en dessous de la fenêtre « Textes sélectionnés »). Il ne reste plus alors qu'à procéder de même pour les autres formes (« Gayssot », « Jean-Claude Gayssot », ...)... puis d'appliquer le recodage souhaité. Attention toutefois, si vous souhaitez recommencer cette opération pour un autre auteur (on peut également procéder de même pour recoder n'importe quel champ des références externes) : n'oubliez pas d'effacer les textes sélectionnés (bouton **Effacer tout**) avant de recoller d'autres textes candidats à un recodage.

2. NAVIGUER DANS LE CORPUS

Il est souvent utile de repérer les usages d'un mot. Soit par exemple l'entité « danger ». Supposons que l'on souhaite aussi obtenir tous les mots proches comme « dangereux », « dangerosité »... L'idée serait de recenser les mots commençant par les lettres d-a-n-g. Un outil est particulièrement adapté à ce besoin.



Comment faire pour retrouver un mot

- Lancer la commande **Outils – Recherche par préfixe / suffixe**
- Taper le début du mot à chercher (ici tapons les quatre lettres dang)
- Cliquer sur le bouton **Par la racine**

On repère les différents mots commençant par « dang » et leur nombre d'occurrences

- Après avoir cliquer sur le mot « dangereux », vous noterez qu'il s'agit d'une qualité. Si ce typage était incorrect, vous pourriez « retyper » ce mot en cliquant sur le bouton droit de la souris, et choisir un autre type. Cette méthode est très pratique pour corriger les erreurs de typage.

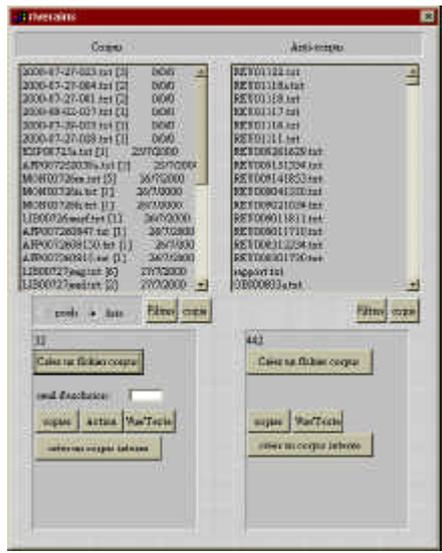


Vous avez sans doute noté que « dangereux » était associé à une catégorie. Nous reviendrons sur cette notion plus bas mais retenez qu'il vous sera possible, par cette fenêtre, d'associer un mot (ou une expression), à une collection, à une catégorie ou à un être fictif.

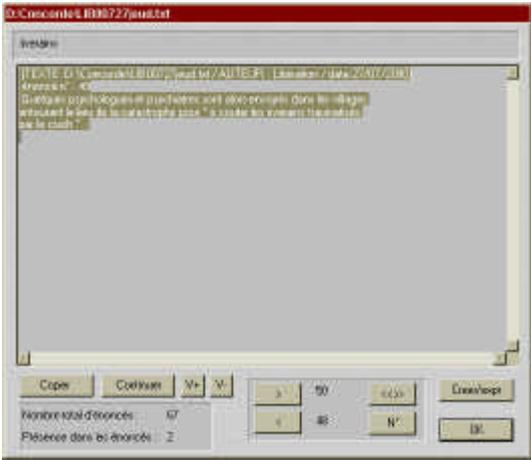
Enfin vous pouvez également rechercher des mots grâce à leur terminaison. Pour repérer, par exemple, tous les termes se finissant par le suffixe (-logie), il suffit de taper des cinq lettres et de cliquer sur le bouton **Par la terminaison**

Remarque : Si vous recherchez les mots commençant par « train », vous trouverez l'expression « train d'atterrissage ». Selon vous, comment se fait-il que cette expression apparaisse dans cette liste ? (pour vous aider, vous pouvez éditer le fichier FR_ETRE.DIC) :

La principale utilisation de la commande **Recherche par préfixe/suffixe** est de repérer les usages d'un mot ou d'une expression construite. Il s'agit donc de recenser les contextes dans lesquelles, tel ou tel mot émerge. Recherchons par exemple le mot « riverains ». Dans mon corpus, je note 56 occurrences. En sélectionnant ce terme et en cliquant sur le bouton **Accès énoncés**, j'obtiens une boîte de dialogue, que vous aviez vue lors de la leçon n°1.



Vous souvenez-vous de la démarche ? Il suffit de cliquer sur l'un des textes de la colonne de gauche puis sur **Action**. Comme au cinéma, vous lancez la commande par laquelle vous vous retrouvez au cœur de l'énoncé ou le mot « riverains » est utilisé (le chiffre entre crochets indique le nombre d'occurrences). Prenez un texte contenant plusieurs fois le mot « riverains ».



Dans l'exemple reproduit ci-dessus, j'ai fait apparaître le texte LIB00727jeud. L'énoncé n°49 contient bien le mot « riverains » est bien présent. Quelques boutons de cette boîte de dialogue vous permettent de circuler dans le texte.

A vous de noter leur effet :

- (pour vous aider, remarquez que le chiffre qui suit est le 50, alors que l'énoncé affiché porte le n°49).....
-
-

N°

V+

V-

Continuer

Vous connaissiez celui-ci : Rappelez sa fonction :

Enfin le bouton **copier** lance la commande bien connue (elle est aussi accessible par le bouton droit de la souris), recopiant le contenu de la sélection dans la presse-papier. Il vous suffit ensuite de **Coller** ce contenu dans votre traitement de texte...

Fermez cette fenêtre. Notez que vous avez la possibilité de trier la liste des textes contenant le mot « riverains », soit par le poids (c'est-à-dire en fonction du nombre d'occurrences de « riverains » dans les différents textes), soit par date (choix activé par défaut). Essayez ! Cette fenêtre propose une autre fonction très importante : le filtre. Il s'agit de repérer, grâce aux références externes, quels sont les auteurs, les statuts, les types de support... (d'une façon générale toutes les variables présentes dans les références externes) qui font le plus usage du mot « riverains ». Appuyez sur le bouton **Filtre**. Une nouvelle boîte de dialogue apparaît :



Ici la fonction de filtrage s'effectue par l'auteur. On pourrait discriminer ces résultats par le type de support. Pour ce faire, cliquer sur **Auteur** puis avec le menu déroulant, choisissez, par exemple, « type de support ». Sur la zone de gauche, l'on repère que c'est la presse nationale quotidienne qui obtient le plus haut score (23 occurrences sur mon corpus). Cela signifie que 23 textes partageant la modalité « Presse nationale quotidienne » (qualité figurant dans les références externes, et saisie de façon automatique grâce au fichier CODAGE.TXT) contiennent le mot « riverains ». Les autres types de support apparaissent par ordre décroissant. Peut-être vous

étonnez-vous de l'absence de libellés sur une ligne. Elle est due à un oubli : les champs des références externes n'ont manifestement pas été remplis pour tous les textes.

Par cette opération, l'on peut donc repérer quels sont les types de support (mais aussi, quels auteurs, quels statut...) utilisant le plus (ou le moins) le mot « riverains ». Il reste un problème. En effet, si dans votre corpus la presse nationale quotidienne est sur-représentée par rapport à d'autres types de support, le nombre d'occurrences aura du coup beaucoup plus de chances d'être important. Pour pallier ce problème, Prospéro est capable de relativiser les valeurs annoncées en fonction du volume occupé par les textes, en ramenant le tout à une base de 100. Cette mesure des écarts relatifs se lance en cliquant sur le bouton droit de votre souris lorsque celle-ci de trouve positionnée sur la zone présentant la liste : choisissez **Mesure des écarts**.



Les règles d'interprétation sont précisées sur la partie droite de la fenêtre. Dans mon cas, la presse nationale quotidienne reçoit une valeur de 153. Cela signifie que si nous prenons comme référence un type de support qui aurait obtenu 100 (la moyenne), le score relatif de la presse nationale est obtenu en multipliant la référence par 1,53. Dans mon corpus, le type « news et magazine » obtient une valeur relative de 667, ce qui signifie que ce type de support parle près de 7 fois plus de « riverains » que les autres. C'est donc une mesure multiplicative qui est ici calculée. Un score relatif de 125 (ou de 80) signifie une augmentation (respectivement une diminution) de 25% par rapport à la valeur de référence (100). Le calcul opéré est détaillé et justifié en annexe.

Pour revenir aux valeurs « brutes », il suffit de choisir « score » dans le menu contextuel (celui qui est accessible par le bouton droit de la souris).

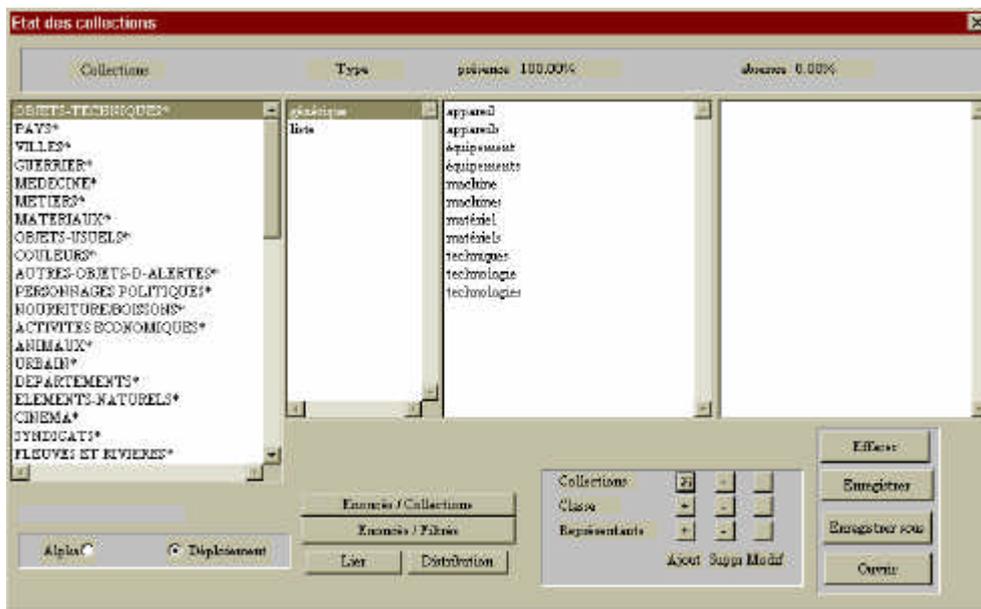
3. LES CONCEPTS

Lors de la leçon n°1, vous avez chargé des fichiers contenant les collections, les êtres fictifs et les catégories. Pour l'instant, nous n'avons toujours pas exploité ces informations que nous désignons par «concepts ». Nous allons apprendre à les définir, à utiliser ce travail. Puis nous nous interrogerons sur ses enjeux épistémologiques et méthodologiques.

A) Les collections

Les textes et les discours ont une capacité à produire des énumérations en puisant dans des répertoires de noms relevant de différentes classifications historiquement constituées. La liste des..., les différentes espèces de..., l'ensemble des..., le répertoire des.... Du coup, étudier le propos des acteurs suppose de recenser les collections qu'ils mobilisent. Il nous faut donc dresser la liste des objets classifiés.

Lancez la commande **Concepts – Collections**. Apparaît la fenêtre suivante.



La liste figurant sur la colonne de gauche recense les différentes collections déjà présentes. Notons que par convention, on a décidé de les inscrire en majuscules, et afin de pouvoir encore mieux les identifier, l'on flanque leur nom d'une étoile (*). Dernière contrainte sur ces libellés : éviter les espaces et les slashes (/). En bas à gauche de la fenêtre, vous avez la possibilité de faire apparaître cette liste soit par ordre alphabétique (« Alpha »), soit par ordre d'importance dans votre corpus (l'option « Déploiement » est activée par défaut).

Il a été prévu un « type » c'est-à-dire un niveau taxinomique permettant de ranger les différents éléments appartenant à une collection donnée. Celui-ci est facultatif. Si vous souhaitez voir

figurer l'ensemble des mots ou expression appartenant à la même collection, et ce sans regroupement supplémentaire, il vous suffira de construire un type général (souvent, on tape trois points de suspension ...) avant de saisir les éléments de la liste. Car, si certaines collections traversent assez bien tous les corpus (la liste des pays, des départements français, des couleurs... restent généralement stables d'un corpus à l'autre), les auteurs de vos textes vont sans doute déployer des collections propres au dossier. Il vous faudra bien sûr créer la collection. Dans le cas de la tragédie du Concorde, l'on pourrait par exemple taper la collection des aéroports, des aviateurs, ou des accidents civils aériens. Avant de s'employer à cette tâche, parcourons les autres informations de cette fenêtre. Les deux colonnes de droite font apparaître la liste des éléments présents et absents dans le corpus. Pour mieux comprendre le fonctionnement de cette boîte de dialogue, cliquez sur la collection OBJETS-TECHNIQUES*. Vous notez que les précédents utilisateurs ont créé deux types : « générique » et « liste ». Le premier contient des mots comme « appareil », « machine », « matériel »... Tous les représentants figurent dans le corpus. L'autre, « liste », rassemble des éléments comme « Airbus », « alimentation électrique », « canalisation »... qui sont présents dans le corpus, mais aussi « barillet », « barre de commande », « benne »... qui eux, sont absents.

D'un point de vue informatique, l'ensemble de ces représentants est enregistré dans un fichier. Ici, il s'agit de ALERTE.COL. Vous aviez indiqué ce document en constituant le projet (fenêtre **Fichier – Gestion du projet**). Avec votre traitement de texte (ou à l'aide d'un éditeur), il vous est possible d'ouvrir ce fichier. Sa structure est la suivante :

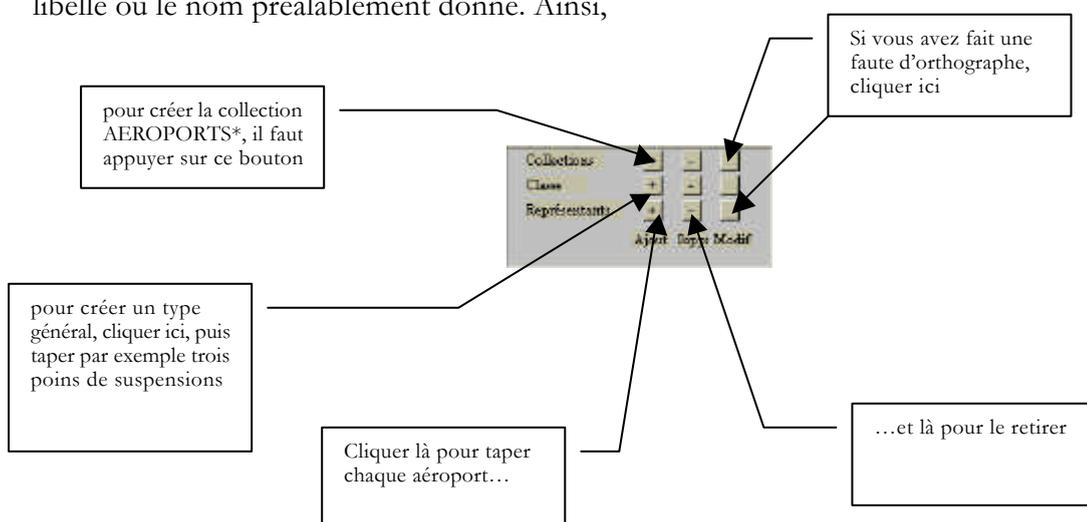
Syntaxe générale	Exemple
Col001	Col001
FICTION	FICTION
NOM DE LA COLLECTION	OBJETS-TECHNIQUES
Nom du type 1	Liste
Élément 1	soudure
Élément 2	condensateur
...	...
Élément n	tuyaux
END	END
Nom du type 2	générique
Élément 1	techniques
Élément 2	machines
...	...
Élément n	tuyaux
END	END
...	...
ENDFICTION	ENDFICTION

Il vous est donc possible créer « à la main », directement avec votre éditeur des collections et d'y ajouter des éléments. Pour être certain de ne pas oublier des aéroports à la collection que vous vous apprêter de constituer, vous pouvez récupérer cette liste sur Internet ou auprès de la DGAC (Direction Générale de l'Aviation Civile) et la coller en respectant la structure indiquée. N'oubliez pas d'enregistrer au format texte. Si vous utilisez la boîte de dialogue ci-dessus, il vous faudra aussi enregistrer. Les boutons **Ouvrir**, **Enregistrer**, **Enregistrer sous**, et **Effacer** (c'est-à-dire fermer) vous permettent d'organiser vos fichiers collections (l'extension est COL) de façon habituelle.

Apprenons à gérer vos collections avec la fenêtre. Un ensemble de 9 petits boutons vous aidera à cette tâche.



Ceux de la première ligne permettent d'organiser les collections, ceux de la seconde, les types (ou classe, ou encore niveau taxinomique) alors que ceux de la dernière ligne concernent les représentants (ou éléments de la collection). La logique est dans tous les cas la même. Le signe « plus » signifie que vous ajoutez quelque chose (une collection, une classe ou un représentant) ; le signe « moins » traduit la suppression de l'objet. Le dernier, quant à lui, permet de modifier le libellé ou le nom préalablement donné. Ainsi,



Pour vérifier les résultats de votre opération, après avoir enregistré, vous pouvez éditer le fichier avec votre traitement de texte. Nous étudierons plus bas comment se distribuent les collections disponibles dans ce dictionnaire.

B) Les êtres-fictifs.

Les Etres-Fictifs ou Personnages servent à expliciter les procédés par lesquels les auteurs - et, par hypothèse, les lecteurs - d'un corpus relient naturellement des entités autour d'un noyau référentiel relativement stable. Créer un Etre-fictif revient ainsi à forger une sorte de macro-acteur ou de personnage géant dont les variations de qualification importent autant que la stabilité. C'est parce que tout le monde admet que « François Mitterrand » et « Tonton » pointent vers le même personnage, que « l'Etat » ou « les pouvoirs publics » renvoient bien au même « acteur », que la création de ces Etres-Fictifs peut être fondée. Il faut néanmoins s'interroger sur la pertinence de la création d'un Etre-Fictif ; notamment il est préférable d'éviter d'avoir recours à un tel concept si sa construction est discutée par les acteurs, si des controverses existent sur ses contours.

Nous discuterons plus tard de ces problèmes ; Attardons-nous pour l'instant à décrire la gestion des êtres-fictifs.

Les manipulations sont très proches de celles pour les collections. Par convention le nom d'un Etre-Fictif est toujours écrit en majuscules et assorti du signe @ (symbole bien connu des internautes mais que l'on utilisait bien avant l'extension de la Toile). On sait ainsi au premier coup d'œil qu'un personnage ou un thème tel que ETAT@, SCIENCE@, AMIANTE@, FAMILLE@ est un Etre-Fictif et qu'il peut donc être déployé.

La liste des représentants des êtres-fictifs est accessible en lançant la commande **Concepts – Etres Fictifs**.

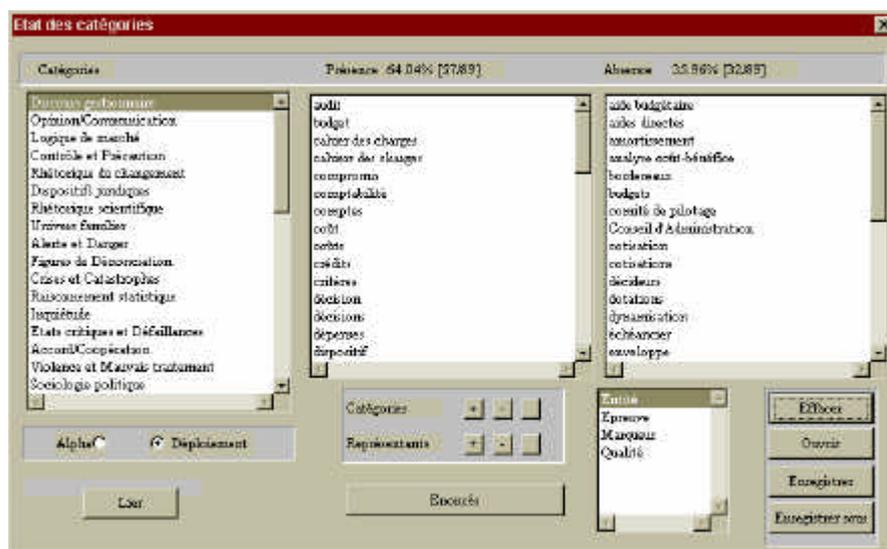


Comme pour la boîte de dialogue permettant de gérer les collections, la liste des êtres-fictifs apparaît sur la colonne de gauche et les types (ou classe) sur celle d'à côté. L'ajout, la suppression ou la modification des noms s'opèrent de la même façon. De même, la syntaxe du fichier rassemblant toutes ces données, dont l'extension est FIC, est semblable à ceux contenant les collections. Editez ALERTE.FIC (avec votre traitement de texte) pour vous en convaincre.

C) Les catégories

Les catégories regroupent des éléments qui, sans être toujours interchangeables (cas limite de la synonymie), entrent dans des rapports de connotation de façon à produire un univers de discours cohérent : s'il y a l'un, l'autre n'est pas loin ou le chemin à parcourir de l'un à l'autre est relativement court. De telle manière que l'on passe plus facilement de Freud à Psychanalyse, Inconscient et Divan que de Freud à Wall Street et à conseiller financier... Contrairement aux Etres-Fictifs, une catégorie ne représente pas un même objet ou ensemble d'objets mais constitue un répertoire organisé autour de figures prototypiques, d'une série de noyaux lexicaux eux-mêmes générateurs, par différenciation intégrative, de satellites (se raccordant au cœur de la catégorie de manière « harmonique »). Ces figures et noyaux peuvent disposer d'une longue histoire – auquel cas ils semblent relativement évidents – ou être précisément en jeu ou en gestation dans le ou les corpus étudiés. Mais, ce qui compte dans la production des catégories c'est avant tout une logique d'opposition : des ensembles de termes tendent à se disjoindre. La construction des différents jeux de catégories n'obéit pas aux mêmes processus cognitifs selon qu'il s'agit des entités, des qualités, des marqueurs ou des épreuves. Du coup, l'on séparera les catégories d'entités, des catégories de qualités,...

La fenêtre, accessible par la commande **Concepts – Catégories**, diffère donc légèrement :



En bas à droite, à côté des boutons permettant de gérer les dictionnaires de catégories (leur extension est CAT), vous trouverez une zone dans laquelle vous décidez sur quelle famille de catégories (entité, épreuve, marqueur ou qualité) vous souhaitez travailler.

La manipulation des boutons est semblable à celle pour les collections et les êtres-fictifs. La structure du fichier informatique (que je vous invite à éditer) tient compte de la famille de catégories. Du coup, vous remarquerez qu'une catégorie d'entités, par exemple, s'identifie par « *ENTITE * » précédant son nom et un « END » suivi de « ENDCAT ».

D) Utilisons les concepts

Construire des collections, des catégories et des êtres-fictifs est une chose. L'intérêt est bien sûr d'étudier comment les auteurs de votre corpus s'appuient sur ces ressources. Vous disposez principalement de deux méthodes pour cela.

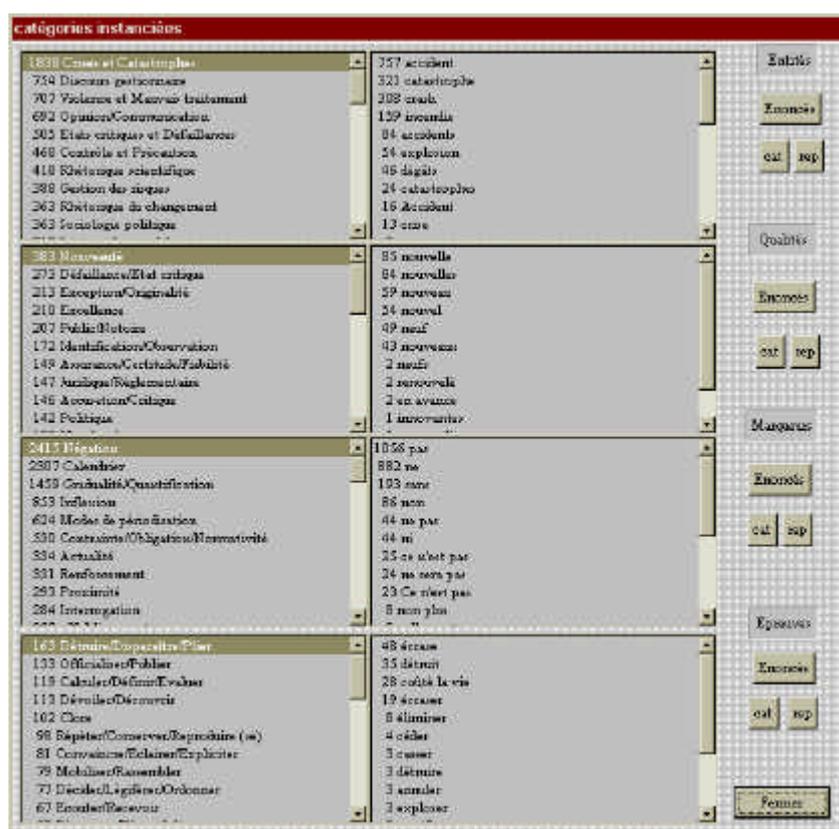
D'une part, le menu **Mode d'accès** vous permet d'accéder facilement aux énoncés dans lesquels ces concepts sont mobilisés. Par exemple, lancez la commande **Mode d'accès – Collections**.



Le bouton **Accès Énoncés** vous ouvrira une fenêtre que vous connaissez bien, puisqu'elle vous permet de zoomer sur un énoncé contenant un élément de cette collection. Du coup, il vous est par exemple possible de recenser les auteurs utilisant le plus une collection. Quel type de support mentionne le plus des HOMMES-POLITIQUES* ?

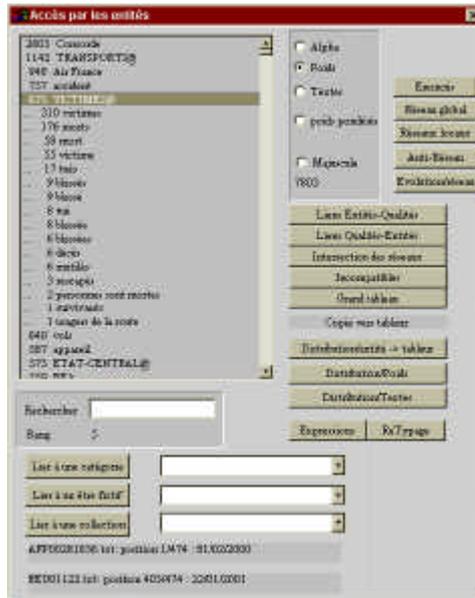
Comment avez-vous fait ?.....

Pour les catégories, la commande **Mode d'accès – Catégories** affichera une fenêtre rassemblant, par ordre décroissant d'importance, les catégories (de toute famille) utilisées par les auteurs. Vous pouvez faire apparaître les textes (et non pas les énoncés) dans lesquelles elles se déploient en cliquant sur le bouton **Textes dans lesquelles la catégorie est dominante**. Il est souvent plus pratique de travailler en dissociant les catégories d'entités, de qualités, d'épreuves et de marqueurs, et surtout d'accéder directement aux énoncés. Pour cela, il vous faut cliquer sur le bouton **Carte-Corpus pour l'ensemble des catégories**.



L'espace se divise horizontalement en quatre zones correspondant respectivement aux catégories d'entités, de qualités, de marqueurs et d'épreuves. Sur la partie gauche, figurent les noms des catégories (rangées par ordre décroissant d'importance dans le corpus). Sur celle de droite, vous trouvez les représentants correspondants. La pression du bouton **Enoncés** lance la liste des énoncés contenant soit la catégorie, soit le représentant selon ce que vous avez sélectionné. Enfin, notez la présence d'un bouton **cat** et d'un autre **rep**. Ils vous permettent respectivement de copier, par exemple dans votre traitement de texte (mais aussi dans votre tableur si vous le souhaitez), la liste des catégories et des représentants. Si le bouton **rep** copie l'intégralité de la liste, celui dédié aux catégories (**cat**) nécessite d'indiquer combien d'éléments vous souhaitez garder.

En revanche, il n'existe pas de commande **Mode d'accès – Etres Fictifs**. Pourquoi ? Parce qu'ils apparaissent directement dans la liste des entités (souvenez-vous **Mode d'accès – Entités**). En effet, si vous avez compris qu'un être-fictif sert essentiellement à désigner un personnage unique, mais dont les modalités de désignation varient, il vous semblera logique qu'il apparaisse avec les autres entités. Le symbole @ vous permettra immédiatement de savoir qu'il s'agit d'une construction contenant donc plusieurs éléments. En double-cliquant dessus, il vous est possible de lister les représentants.



Vous pouvez également accéder aux énoncés contenant les collections, catégories et êtres-fictifs en utilisant les boutons des fenêtres permettant de construire ces concepts. La boîte de dialogue **Concepts – Collections**, par exemple contenait quelques boutons que nous n'avions pas recensés.



Essayez les deux premiers, et notez l'effet obtenu :

Enoncés / Collections

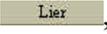
.....

Enoncés / Filtrés

.....

~~Essayez à présent celui là :~~ **Distribution** La fonction provoque des sorties intempêtes du programme

.....

Le bouton , quant à lui, s'utilise différemment. Il s'agit de la dernière étape permettant d'associer un élément, apparaissant lors de vos investigations, à une collection (mais la logique est la même pour les êtres-fictifs et les catégories). Ainsi, lorsque vous parcourez la liste des entités (respectivement des qualités, des épreuves...) vous disposez de boutons permettant de lier directement l'entité courante à un concept. De même lorsque vous recensez un mot ou une expression par **Outils – Recherche par préfixe/suffixe**, le menu contextuel (celui obtenu par le bouton droit de la souris) vous permet d'accéder à la même fonctionnalité. Une fois cette commande lancée, le bouton  vous permet de valider l'association.

E) Remarques sur les usages des concepts

Dès lors que l'on vise la compréhension de processus ou de dossiers longs et tortueux le chercheur doit engager des hypothèses sur les « structures » et les « transformations » en jeu dans le corpus. C'est précisément le travail réalisé lorsque l'on utilise des concepts. D'aucuns pensent qu'en ce faisant, en introduisant de l'interprétation, on s'éloigne de l'objectivité scientifique. Nous pensons qu'ils se trompent. En effet, cette démarche apporte un surcroît d'objectivité dans la mesure où il contraint à l'explicitation des catégories et des cadres sémantiques dans lesquels opère l'utilisateur et qui, à défaut de pouvoir être tracés et rendus visibles dans le système, resteraient tacites ou invisibles. En outre, certains objets ou phénomènes donnent lieu à une pluralité d'interprétations possibles. Il ne s'agit pas d'une aporie liée à une démarche perspectiviste mais bien d'une propriété des objets ou des phénomènes. On ne « lit » pas de la même manière le dossier « Concorde » selon que l'on s'intéresse aux enjeux économiques, à l'évolution des technologies, aux instances d'expertise ou de contrôle, aux mouvements de protestation ou aux controverses sur les risques.

La production des concepts vise ainsi à créer un espace intermédiaire permettant la rencontre entre des angles d'attaque cohérents du point de vue de l'interprète et des saillances ou des dispositions identifiables dans les textes eux-mêmes.

Cependant souvent, le prospérien débutant a du mal à identifier les codages pertinents. A cette difficulté, s'ajoute parfois celle concernant le type de concept (collection, être-fictif ou catégories) qu'il doit construire. Ces problèmes sont liés et peuvent être résolus en se posant quelques questions simples.

Si vous souhaitez rassembler sous un même concept plusieurs éléments, il est à mon sens salutaire de s'interroger sur les résistances à ce regroupement. Serait-il facile d'obtenir un consensus pour désigner par un seul terme, les riverains ? On voit bien que tout le monde (tant les chercheurs que les acteurs) font allusion à la même idée lorsqu'il parlent « d'habitants proches », « des population des environs » ou « des gens aux alentours ». Or cette idée renvoie à une seule entité. On peut la nommer « les riverains » qui du coup, invite à créer un être fictif.

Si l'on s'intéresse maintenant à un autre ensemble qui, cette fois, renvoie à des entités différentes (et non, comme précédemment, à une même idée décrite de plusieurs façons), l'on peut aussi se demander dans quelle mesure l'on obtiendrait un accord unanime pour regrouper ces éléments. Si tel est le cas, il faut créer une collection.

En revanche, construire les catégories suppose un travail d'interprétation de la part du chercheur. Il est clair que les catégories, précisément parce qu'elles produisent des rapprochements conceptuels entre des éléments qui peuvent être matériellement séparés (dans le corpus), incorporent la plus grande part de projection de la part de l'utilisateur (celui dont l'épistémologie de base dit qu'il met des lunettes pour « voir la réalité »). Mais là encore c'est bien le sens commun que l'on cherche à authentifier ou pour le moins à clarifier. A ce titre, il faut essentiellement travailler sur la langue. La relecture (ou la lecture) des travaux de linguistes, notamment de O. Ducrot, constitue un moment incontournable.

4. LA DEMARCHE DU PROSPERIEN

Votre attention a principalement été retenue, jusqu'à présent, par les différentes options des boîtes de dialogue, par les boutons, par des procédures. Vous avez compris l'utilité de chaque commande mais vous ne voyez sans doute pas trop comment s'organise le travail du chercheur avec cet outil.

Il faut avant tout avoir une bonne connaissance du corpus. Le lire et le relire en tentant de trouver des marqueurs dans la langue. L'une des priorités est de créer l'ensemble des expressions. En parcourant la liste des entités, l'on s'interrogera sur la présence de « bureau ». Du coup, on « descendra » dans les énoncés pour remarquer qu'il s'agit bien sûr du «bureau enquête accident ». Puis on construira alors l'expression. Pour savoir si d'autres formes existent, on peut alors chercher tous les «Bureau ». En ouvrant les énoncés, on s'aperçoit que l'on a « Bureau Enquêtes Accident » (avec des majuscules) mais aussi « Bureau Enquêtes-Accident », puis tout un

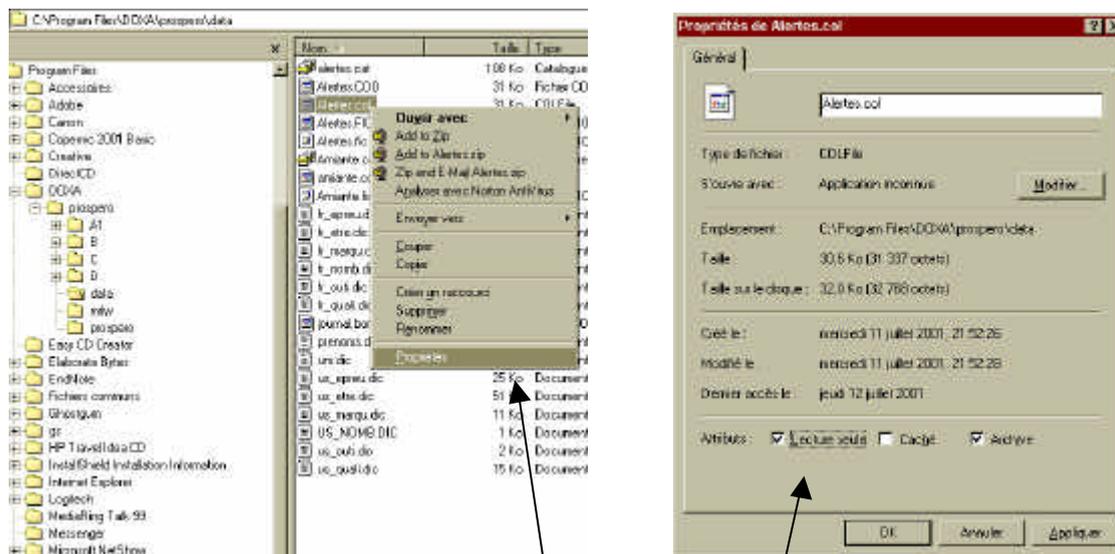
tas de variations liés à la mise au pluriel de l'un de ces trois termes. Du coup, il faut construire toutes ces expressions et créer un être-fictif les rassemblant.

De même, en parcourant les textes on peut avoir l'idée d'une collection regroupant tous les aéroports. Mais lorsque l'on nomme une ville, s'agit-il de son aéroport ? Pas forcément car on pourrait imaginer des expressions comme «les habitants de la ville de Roissy » ou « comme le maire d'Orly, celui de Roissy demanda... ». Avant de coller rapidement Roissy et Orly dans cette liste comprenant par ailleurs des noms posant moins de problème comme «Lyon-Satolas » ou « Roissy-Charles-de-Gaulle », il faut prendre la peine de retrouver les énoncés pour s'assurer que Roissy ou Orly renvoient bien à l'aéroport. Si tel n'était pas le cas, il faudrait trouver un marqueur pour isoler l'expression « la ville de Roissy » ou « l'aéroport de Roissy »... et créer l'expression.

Du coup, le travail avec Prospéro ne peut se réduire à un premier moment théorique où l'on construirait expressions, collections et autres concepts, et un second où l'on exploiterait le travail en dressant des listes et des tableaux. En fait, s'il est souhaitable de travailler au plus vite avec des expressions créées, travailler avec Prospéro revient le plus souvent à un va-et-vient permanent. On a une idée, on cherche les usages d'un mot, on note qu'il est souvent employé avec d'autres au point d'envisager la création d'une expression ou d'un concept, on vérifie en parcourant les énoncés que la formule est suffisamment « pure » pour être isolée, puis après avoir construit cette forme, on recense le nombre d'occurrences et les auteurs. En parcourant les énoncés pour vérifier notre hypothèse, on retombe sur une figure rhétorique intéressante qui appelle un nouveau travail du même type...

Enfin, il convient de s'arrêter à la gestion des dictionnaires de concepts. Très souvent, on hérite d'un ensemble de collections, catégories et d'êtres fictifs. Vous avez sans doute compris que la pertinence de votre analyse tient à votre capacité à interroger ces listes constituées par de précédents utilisateurs. Car d'une part, il est des concepts propres au dossier que vous étudiez (la liste des produits dopants, les formules langagières des agents de conduites, les sigles utilisés dans un procès...). Mais d'autre part, des représentants méritent d'être débranchés dans certains corpus. Le mot « piste », souvent associé à une catégorie « perspective » pollue singulièrement ce concept dès lors que l'on s'intéresse à un aéroport. De même, travailler sur des débats sur la forme de kayak de course suppose de « débrancher » la collection MOYENS-DE-TRANSPORT* ce sans quoi toute allusion au « bateau » contribue à augmenter le score de ce concept. C'est la raison pour laquelle à chaque projet correspond un jeu de concept qui lui est

propre : c'est pourquoi le cadre d'analyse (c'est-à-dire ces dictionnaires de concept) se définit lors de la création du projet. Pour vous y retrouver, le plus simple est, à mon sens, de laisser quelques dictionnaires de concepts standard à un endroit de votre disque dur (pourquoi pas à côté du programme prospéro) mais de les flanquer de l'option « lecture seule ». Je vous rappelle que ce choix peut être indiqué lorsque vous travaillez avec l'explorateur Windows, en activant la commande **Propriétés** accessible avec le bouton droit de votre souris.



Avec le bouton droit de la souris, choisissez Propriétés... puis **Lecture seule**

Ainsi, vous ne pourrez pas enregistrer vos modifications sous ce même fichier (qui restera donc toujours identique, comme une base de travail à laquelle on peut toujours revenir). Vous serez obligé de sauvegarder vos collections (respectivement vos êtres-fictifs et catégories) sous un autre nom ou sous un autre endroit. Le plus simple est sans doute d'enregistrer vos concepts à l'endroit même où se trouvent les textes et votre projet, en donnant le même nom. Ainsi, je vous propose d'enregistrer les modifications réalisées dans les collections (respectivement les catégories et les êtres fictifs) pour le dossier « concorde » dans le répertoire contenant le CONCORDE.PRC sous le nom de CONCORDE.COL (respectivement CONCORDE.CAT et CONCORDE.FIC).

5. EXERCICES

Exercice 1 :

Dresser la liste de tous les pays mentionnés dans le corpus. Quels sont les auteurs ayant le plus recours à cette liste ?

Exercice 2 :

Créer un être fictif BEA@ (Bureau Enquêtes Accidents). Etablir la liste de tous les représentants avec leur nombre d'occurrences.

Exercice 3 :

Retrouver quelques énoncés significatifs dans lesquels il est question d'accusation.

Exercice 4 :

Etablir la liste de toute les « sciences » mentionnées dans le corpus (travaillons, pour simplifier, la terminaison « -logie ». En faire une collection.

Exercice 5 :

Supposons que vous souhaitiez distinguer les textes en deux grandes catégories : l'une rassemblant toutes les productions de journalistes (les dépêches, les quotidiens, les magazines...), l'autre le reste. Utiliser les références externes pour marquer cette distinction. Est-elle intéressante pour étudier des êtres-fictifs comme les VICTIMES@, LES-RIVERAINS@, des catégories comme les « figures de dénonciations » ou « l'actualité » ou encore une collection comme les HOMMES-POLITIQUES* ?

Exercice 6 :

Vous travaillez sur le dopage. Après avoir récupéré la liste des produits dopants, est-il préférable selon vous de créer une collection, une catégorie ou un être fictif? Pourquoi? Donnez des exemples pour les deux concepts que vous n'avez pas retenus.

Leçon n°3 : Je repère des propriétés du corpus

CE QUE VOUS ALLEZ APPRENDRE :

- *Repérer les personnages et leurs rencontres*
- *Repérer ce qui est mesuré dans un corpus*
- *Repérer des expressions particulières grâce aux formules*
- *Identifier des répertoires, des régimes discursifs, des jeux d'acteurs, des configurations et des figures*

1. LES PERSONNAGES

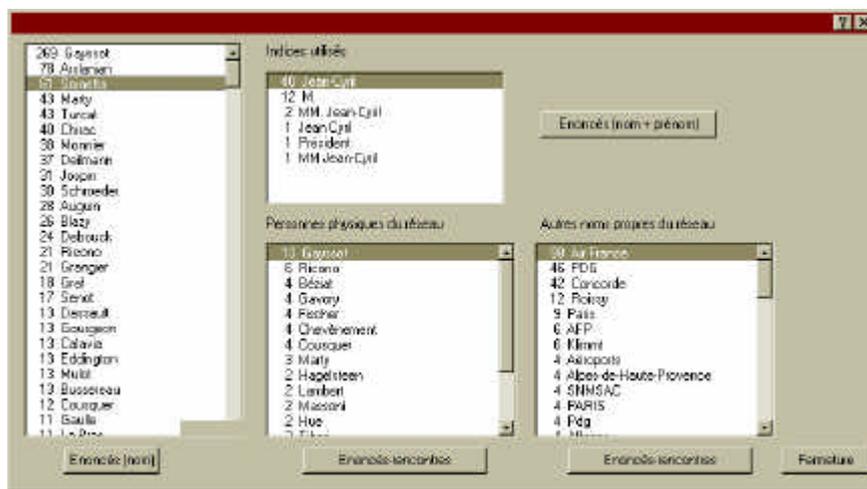
Pour identifier les noms de personnes présents dans le corpus, Prospéro s'appuie sur quelques marqueurs utilisant notamment la liste des prénoms (son adresse est indiquée dans la boîte de dialogue **Fichier – Configuration du projet**).

En lançant la commande **Mode d'accès – Listes – Patronymes – Prénoms**, vous accédez à une fenêtre contenant, sur sa partie gauche, une colonne rassemblant tous les prénoms saisis à ce jour. Vous pouvez bien sûr en rajouter (éventuellement en supprimer en cas de faute de frappe) si votre corpus contient des personnages aux prénoms exotiques non présents dans cette liste : les boutons **Ajouter** et **Supprimer** sont destinés à ces opérations. En cas d'ajout, il vous faudra penser à enregistrer cette information dans le fichier PRENOM.DIC (bouton **Enregistrer**). Si la liste n'apparaît pas sur la colonne de gauche, cela signifie sans doute que le PRENOM.DIC ne figure pas à l'adresse indiquée dans la fenêtre **Fichier – Configuration du projet**. En cliquant sur **Ouvrir** vous pouvez indiquer l'endroit où se situe ce document mais cette opération sera à renouveler à chaque ouverture d'un projet. Il semble donc préférable d'indiquer une fois pour toute où se situe le fichier des prénoms dans la configuration du projet.

Lancer le **calcul** en pressant le bouton approprié. La colonne du milieu se remplit. En cliquant sur un des prénoms, vous obtenez, sur la partie droite, les noms propres suivant ce prénom. Ce qui peut apparaître comme un petit jeu (*dites-moi qui, dans le corpus, se prénomme Alain ?*) est en fait une opération très importante car désormais, Prospéro dispose de l'ensemble des marqueurs précédents un nom patronymique (les « Monsieur », « Madame »,... et les prénoms).



Vous pouvez donc fermer cette fenêtre... et ouvrir celle-ci : **Mode d'accès – Listes – Patronymes – Noms.**



Elle fait apparaître la liste des noms patronymiques (sur la colonne de gauche) grâce aux indices apparaissant à droite. Ainsi, en cliquant sur un personnage, on peut repérer comment il est « introduit ». Le bouton **Enoncés (nom + prénom)** permet d'accéder à la liste des énoncés dans lesquels l'individu est mentionné sous la forme sélectionnée. Du coup, on peut repérer quel type de support, quel auteur,... préfère utiliser telle ou telle forme pour désigner les personnages, ce

qui peut être une piste intéressante. Car il est des indices suggérant une dénonciation. Lorsqu'un personnage se voulant au service du peuple se voit flanqué d'un « Monseigneur », ou qu'un président d'une institution, fier de son titre, est introduit par son prénom ou un simple « M. », l'on peut considérer qu'il s'agit d'une figure de dénonciation dont on peut étudier la source, les causes, ... Situé en bas à gauche, le bouton **Énoncé (nom)** quant à lui permet un accès direct à l'ensemble des énoncés indépendamment des indices introduisant le personnage.

Une partie de la fenêtre fait apparaître les « personnes physiques du réseau ». Le mode de calcul permettant de déterminer le réseau sera précisé dans la prochaine leçon. Retenons pour l'instant que le nombre indiqué traduit la force du lien entre les deux personnages. Ainsi, l'illustration ci-dessus montre que le nom de Spinetta apparaît plus souvent et de façon plus proche avec celui de Gayssot qu'avec celui de Ricono ou de Béziat. Le bouton **Énoncés - rencontres** ouvre une fenêtre qui vous est désormais familière ; il vous est donc possible de zoomer sur les énoncés rassemblant les noms des deux personnes.

La démarche est la même pour les autres noms propres du réseau. Un nom propre, pour Prospéro est une entité commençant par une majuscule. Cette remarque explique le « bruit » apparaissant dans la colonne de droite. Cette fenêtre ainsi que le bouton **Énoncés – rencontres** permettent de repérer les liens entre des individus et des institutions.

Qu'il s'agisse de n'importe laquelle des trois petites fenêtres (indices, rencontre avec des personnes physiques, ou avec d'autres noms propres), il vous est possible d'utiliser le bouton droit de la souris, pour copier ces listes vers le traitement de texte.

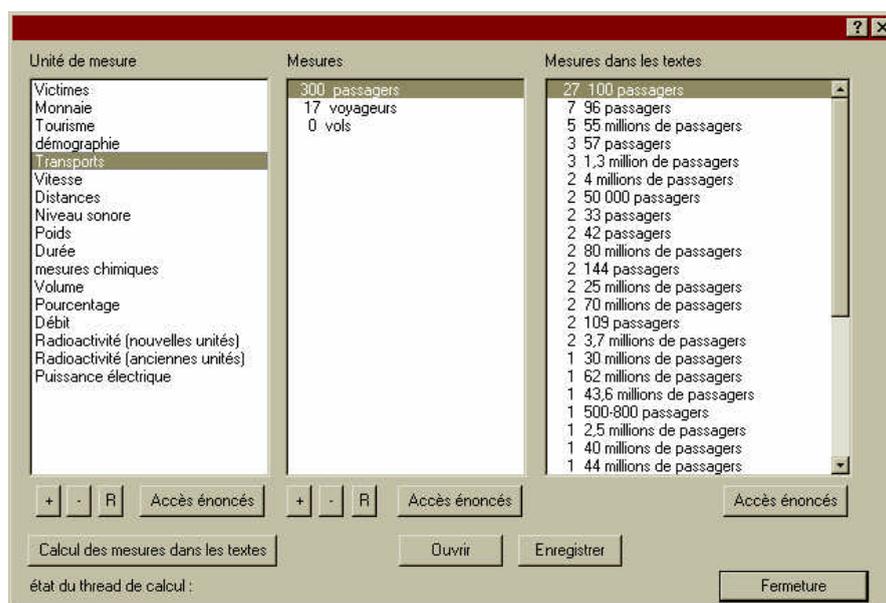
2. LA MESURE DANS LE CORPUS

Recenser ce qui est mesuré dans vos textes participe de la description et de l'analyse du corpus. La métrologie constitue en effet une ressource des acteurs.

Un fichier, nommé le plus souvent UM.DIC (enregistré au format texte) contient la liste des unités de mesure. Comme pour PRENOM.DIC, son adresse doit être précisée dans la fenêtre **Fichier – Configuration du projet**

La commande **Mode d'accès – Unité de mesure** ouvre une boîte de dialogue faisant apparaître sur une colonne de gauche, ce qui est compté. Les boutons , situés en dessous, permettent

respectivement de rajouter, de supprimer ou de renommer des éléments de cette liste. La colonne du milieu rassemble les unités de mesure pour la « grandeur » sélectionnée. Les boutons **+** **-** **R** fonctionnent de même. Les modifications effectuées sur ces listes doivent être enregistrées grâce au bouton **Enregistrer**. On peut ouvrir un fichier d'unités de mesure ou le faire apparaître (si toutefois, l'adresse mentionnée dans la fenêtre **Fichier – Configuration du projet** était erronée) grâce au bouton **Ouvrir**. Comme pour les prénoms, on lance le calcul en pressant le bouton approprié : **Calcul des mesures dans les textes**.



La colonne de droite recense les nombres figurant devant l'unité de mesure sélectionnée (si elle reste vide, cela signifie que l'on parle, par exemple de « blessés », sans les dénombrer précisément). Les trois boutons **Accès énoncés** permettent d'accéder aux énoncés dans lesquels apparaissent respectivement la métrologie (la monnaie, le volume, le débit...), l'unité de mesure utilisée et les valeurs repérées selon cette unité.

Cet outil permet donc de décrire :

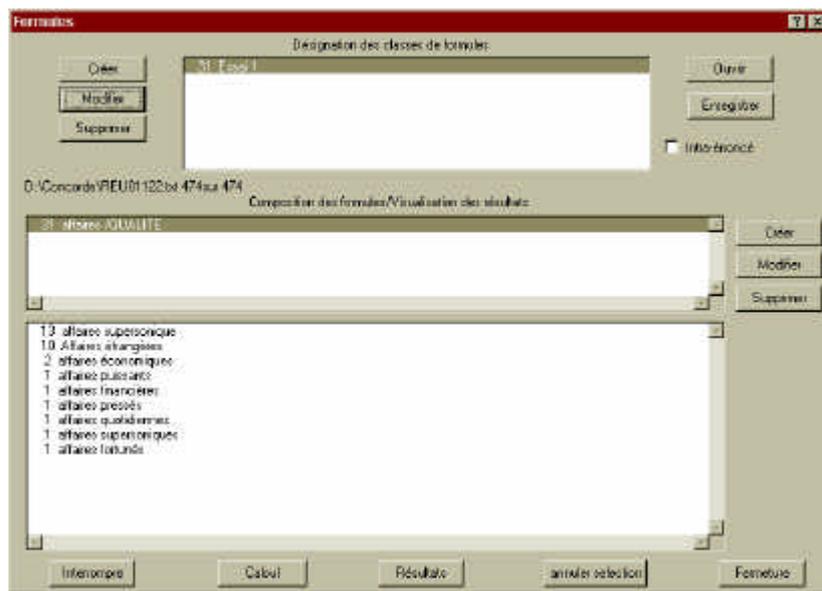
- les ordres de grandeurs (dans tel corpus, on parle en millions de dollars lorsque dans tel autre il n'est question que de milliers de Francs),
- les valeurs revenant souvent indiquant l'absence de controverse (le nombre de morts liés au crash, la capacité d'un avion,...)
- celles dont la variation indique une possible controverse (et ouvre donc une piste pour le chercheur)
- ce qui est compté (et ce qui ne l'est pas) et les métrologies mobilisées par les acteurs.

Sur ce dernier point, il est sans doute utile de préciser que l'on peut disposer de plusieurs fichiers d'unités de mesure car si le gigabecquerel renvoie toujours à la radioactivité, une « unité » comme « cas » en revanche peut tout aussi bien servir à comptabiliser des victimes (dans les crises sanitaires) que des affaires, des coureurs dopés,... Du coup, comme dans les concepts (catégories, collections et êtres-fictifs), un mot, une expression peut figurer dans un dictionnaire pour l'étude d'un corpus, mais pas dans un autre...

3. LES FORMULES

Vous avez appris à retrouver rapidement un mot, une expression dans votre corpus. La commande **Outils – Recherche par Préfixe / suffixe**. Cette commande, pratique, vous permet notamment de visualiser rapidement le « score » d'un mot ou d'un groupe de mots (puisque la recherche « par la racine » revient à une lématisation). Mais vous avez surtout compris son utilité lorsque vous recherchez à construire des expressions, comme « affaires étrangères ». Il serait donc pratique de lister les adjectifs suivant le mot « affaires », d'une part pour créer les expressions nécessaires et d'autre part pour recenser les qualités reçues par cette entité. Cette opération est très facile dès lors que l'on connaît les formules.

Lancez la commande **Inférences – Formules**. La boîte de dialogue comprend trois parties. En haut, vous entrez une classe de formule, c'est-à-dire un nom sous lequel vous pourrez enregistrer plusieurs formules. La zone centrale vous sert à entrer les formules alors que celle située en bas fait apparaître les résultats.



Dans la partie « classe de formules », cliquez sur **Créer** puis tapez, par exemple, Essai 1. Les boutons **Modifier**, **Supprimer**, **Enregistrer** et **Ouvrir** servent, comme leur nom l'indique, respectivement à renommer une classe de formules, à la supprimer, à la sauvegarder (l'extension est FRM) ou à la récupérer.

Nous allons à présent saisir votre première formule. Cliquez sur le bouton **Créer** situé sur la droite. Tapez :

affaires /QUALITE

... en respectant l'espace après « affaire ». En revanche, il ne faut pas en mettre devant QUALITE qui doit obligatoirement être tapé en majuscule. /QUALITE est un élément du langage des formules. Comme tous ceux que nous allons étudier, il commence par une barre oblique (/) et est suivi d'un mot-clé écrit en majuscule.

Validez, puis cliquez sur le bouton **Calcul**, en bas de la fenêtre. Prospéro parcourt l'ensemble des textes de votre corpus en recherchant les mots « affaires » (sans se soucier de la casse, c'est-à-dire sans tenir compte de la présence des majuscules) suivis d'une qualité. Il utilise pour cela le typage expliqué lors de la leçon 1. Cela signifie qu'il ne considèrera que les mots (ou expressions) typés comme une qualité.

En cliquant sur **Résultats**, vous voyez apparaître la liste... (vous pouvez aussi double-cliquer sur la formule). En sélectionnant par exemple « Affaires étrangères », vous pouvez utiliser le bouton droit de la souris pour ouvrir les énoncés, créer l'expression, la lier à un concept, copier la liste...

Il est temps d'apprendre d'autres mots-clés vous permettant de construire de nouvelles formules. Comme il existe /QUALITE, pour désigner n'importe quelle qualité recensée comme telle par Prospéro, nous disposons aussi de /ENTITE, /EPREUVE, /MARQUEUR.

Ainsi, peut-on par exemple chercher les entités qualifiées, qui très souvent, invitent à la création d'expressions. Pour cela il suffit de créer une deuxième formule (dans la même classe) puis de taper :

/ENTITE /QUALITE

L'on peut aussi rechercher les entités doublement qualifiées comme « informations infondées et erronées ».... Pour ce, il faut entrer la formule :

/ENTITE /QUALITE et /QUALITE

On peut recenser les avocats intervenant dans le dossier. Le mot-clé /MAJ désigne un mot commençant par une majuscule. Du coup des formules comme :

Me /MAJ ou Maître /MAJ

... vous permettrons de créer rapidement la collection des avocats.

Comment ?.....
.....

On peut également chercher ce que font les acteurs. Pour cela, il faut savoir /PERS désigne une personne. Mais, pour l'utiliser, il faut avoir préalablement lancé le calcul des personnes (cf. 2.) Dans ces conditions, on peut écrire une formule comme

/PERS /EPREUVE

... ce qui nous permet de noter que nos personnages s'emploient à arriver, à décider, à déclarer...

Il reste que cette dernière formule ne permet d'attraper des expressions comme « Machin, énervé, répondit... ». Car la qualité (ici, « énervé ») ou le marqueur (« sereinement », « calmement »,...) empêche à la dernière formule de s'arrêter sur ces figures. C'est la raison pour laquelle, il a été créé la notion de « trou »... En modifiant la précédente formule (le bouton **Modifier** placé sur la droite sert précisément à cela), de la façon suivante :

/PERS /T=3 /EPREUVE

... permettra de chercher un personnage suivi d'une épreuve mais éventuellement entrecoupé de trois éléments (une virgule, un marqueur, une qualité,...). Notons qu'il gardera néanmoins les expressions du type « Machin déclara... »

Parfois, Prospéro s'arrête sur une telle série qui déborde sur l'énoncé suivant. Par exemple, vous pouvez récupérer des extraits comme celui-ci : « ... Chirac. Il a annoncé... ». Pour éviter ceci, vous pouvez exiger que Prospéro ne déborde pas sur l'énoncé suivant en cochant la case **intra-énoncé**.

A la notion de « trou » s'exprimant par le mot clé /T, correspond celle de « plein » : /P. Il peut être précieux de recenser les usages d'un mot et de repérer automatiquement le ou les éléments le précédant (ou le suivant). En construisant une formule comme :

/P=2 sécurité

... vous obtenez des expressions comme « mesure de sécurité », « niveau de sécurité », « règle de sécurité »... De même, si vous vous intéressez aux principes guidant vos acteurs, il vous suffit d'écrire :

au nom de /P=3

Vous récupérez ainsi des expressions comme « au nom de l'intérêt général » (ou d'autres expressions renvoyant à des principes moraux supérieurs) mais aussi les délégations comme « au nom de la compagnie », « au nom de ce comité »...

Il est possible de ne chercher qu'une partie d'un mot (au sens de Prospéro, il peut s'agir d'expressions). Si vous ne souhaitez que le début, utilisez /COM= (pour « Commencer»). Ainsi /COM=décid vous renverra toutes les formes conjuguées du verbe « décider ». Symétriquement, /TERM= vous permettra d'obtenir tous les mots se terminant par une forme fixée. /TERM=isme permet souvent de vérifier le niveau d'idéologie dans le corpus (capitalisme, marxisme, sociologisme, sexisme...), par la sélection des mots s'achevant en *-isme*. Enfin, si vous voulez repérer une partie des mots ou expressions qui soit ni le début, ni la fin, utilisez /RAC=. Ainsi /RAC=prend permettra de recenser des mots comme « éprendre », « surprendre », « comprendre »... mais pas « prendre ». En effet, un caractère au moins doit précéder la chaîne recherchée. Essayez de deviner ce que pourraient renvoyer les formules suivantes... et vérifier.

/PERS /T=3 /COM=décid /P=4

Exemple :
.....

/ENTITE /T=4 /TERM=aient

Exemple :
.....

/ENTITE /RAC=trouv

Exemple :
.....

/RAC=métal /T=3 /EPREUVE

Exemple :
.....

Comment se fait-il, selon vous que « lamelle métallique » apparaisse alors que /RAC ne retient normalement pas les chaînes de caractères commençant pas métal ?

.....
.....

Si l'on cherche les formes conjuguées d'un verbe, on peut aussi le mot-clé /CJ=. Deux contraintes, pour cela : d'une part le verbe doit figurer parmi les catégories d'épreuve et, d'autre part, il doit être du premier groupe. Ainsi, on peut écrire

/PERS /T=5 /CJ='Lancer une alerte'

Vous disposez enfin de deux autres mots-clés permettant de rechercher des mots dans votre corpus.

- /NBRE permet de s'arrêter sur n'importe quel nombre (au sens de ceux définis dans les dictionnaires de Prospéro, c'est-à-dire aussi bien « 123 » que « trois »). Par exemple, une formule comme /NBRE /P=2 vous permettra de compléter votre liste des unités de mesure (pieds, corps, °C,...). Comment procéder ?

- /MO désigne n'importe quel mot-outil (« de », « à », « le »...). Du coup, en tapant une formule comme /ENTITE /MO /ENTITE, vous pouvez attraper quelques expressions non encore construites comme « Aéroport de Paris », « tour de contrôle », « vitesse de croisière »,...

Mais ce n'est pas tout. Les formules vous permettent également d'utiliser vos concepts (être-fictif, collection, catégorie).

- Vous pouvez notamment construire une formule en référence à un être-fictif grâce au mot-clé : /EF=. Ainsi dans une formule comme /EF=PREFET@ /T=5 /EPREUVE recherche tous les extraits correspondant à ce que fait, la préfecture, le préfet... ou n'importe lequel des représentants de l'être-fictif PREFET@.

- Il est de même possible de faire référence à une collection. Le mot-clé /COL= permet en effet de désigner n'importe quelle collection. Pour identifier les actions des différents professionnels du dossier, vous pouvez par exemple construire une formule comme /COL=METIERS* /T=5 /EPREUVE.

- Les catégories se répartissent en quatre familles : les catégories d'entités, de qualités, de marqueurs et d'épreuves. Du coup, faire référence dans une formule à une catégorie suppose de préciser son type. Il existe donc quatre mots-clés pour les désigner : /CAT_ENTITE, /CAT_QUALITE, /CAT_MARQUEUR et /CAT_EPREUVE.

Deux remarques sur la syntaxe de ce type de formules. D'une part, les noms de vos catégories de doivent pas comporter de barre oblique (/). D'autre part, si ces noms comportent plusieurs mots, l'ensemble doit être mis entre guillemets anglais (l'apostrophe). Par exemple, vous devez écrire :

/ENTITE /CAT_QUALITE='Défaillance-Etat critique'

ou encore :

/EF=RIVERAINS@ /T=5 /CAT_EPREUVE='Accuser-Critiquer-Se plaindre'

D'une façon plus générale, si vous souhaitez mentionner une expression contenant une apostrophe, vous devez la faire précéder d'un anti-slash (c'est la barre oblique dans l'autre sens « \ » que l'on obtient en appuyant simultanément sur les touches **Alt Gr** et 8. Par exemple, vous pouvez construire la formule : /ENTITE /T=15 'de l'ordre de' /NBRE /P=3.

Mais il est aussi possible de simplifier l'écriture de certaines formules en utilisant les classes de formule. Le principe est le suivant : vous identifiez un ensemble de formules que vous regroupez dans une classe de formules. Par exemple, vous pouvez créer une classe de formules (en haut de la fenêtre **Inférences – Formules** que vous appelez « principes ». Vous construisez des expressions du type au nom de /P=5 ou en vertu de /P=5....

Dans un autre groupe de formules, vous pouvez utiliser cette classe de formules avec le mot-clé /CF=. Par exemple, on peut écrire :

/PERS /T=10 /CF=principe ou /P=5 /EPREUVE /CF=principe

Il est donc clair que les formules permettent non seulement de construire des expressions mais aussi d'identifier des rhétoriques et des expressions singulières, de repérer de façon exhaustive ce que font les acteurs et ce qu'ils explicitent, de recenser leurs usages de mots ou d'expressions marquant des points d'inflexion dans leurs descriptions du dossier (par exemple, « désormais », « de plus en plus », « à un moment donné »), et de lister les ressources mobilisées (notamment les collections) dans leurs argumentations.

Pour tirer le plus grand bénéfice de ce puissant outil, il convient de bien organiser les formules. Je suggère de rassembler les formules que l'on peut exécuter au début de l'analyse de chaque corpus. Par exemple on peut enregistrer sous le nom de STANDARD.FRM plusieurs classes de formules. L'une recenserait tous les avocats, médecins, et autres individus identifiables par leur titre : elle permet de construire rapidement les collections d'individus intervenant sur le dossier. Une seconde aurait vocation à faciliter la création d'expressions non encore construites (avec des formules du type /ENTITE de /ENTITE, les surqualifications...). Puis, pourraient figurer dans ce fichier une série de formules visant à repérer ce qui est explicité (repérer ce qui précède l'usage de

« c'est-à-dire » ou d'autres marqueurs de la catégorie « explicitation »), une autre ayant vocation à recenser ce que les acteurs font, ce qu'ils disent...

Après avoir éprouvé ces formules sur plusieurs corpus, il vous suffit de lancer l'ensemble lorsque vous abordez un nouveau dossier. Vous pourrez dans un autre fichier FRM, construire des formules propres à vos hypothèses, aux questions que vous vous posez, et aux spécificités du projet.

4. REPERER LES CONFIGURATIONS

Analyser le discours des acteurs c'est recenser les ressources qu'ils mobilisent. Aussi, est-il précieux de pouvoir repérer des textes typiques dans lesquels apparaissent soit un ensemble de collections fixées par le chercheur (nous parlerons de répertoires), soit des listes de personnages ou d'entités qui se croisent (que nous avons appelées des jeux d'acteurs), soit un ensemble de liens stabilisés entre des catégories (il s'agit des régimes discursifs), soit enfin, un ensemble de liens (ou d'absence de liens) établis par l'utilisateur entre des thèmes ou des personnages, des catégories ou des collections. Il s'agit alors d'un niveau de synthèse au-dessus des répertoires, des régimes discursifs et des jeux d'acteurs : les configurations.

1) Les répertoires

Admettons que nous souhaitons identifier les textes dans lesquels on parle surtout de techniques et accessoirement de l'activité économique.

Lancez la commande **Inférences – Répertoires**. La boîte de dialogue invite à **créer** un répertoire. Appuyez sur le bouton approprié, puis donnez un nom (comme « Exemple de répertoire »). Il nous faut alors le définir : pour ce faire appuyer sur le bouton **Editer**.



Le principe est le suivant. Vous choisissez les collections que vous souhaitez voir apparaître les sélectionnant puis en appuyant sur la flèche : . Vous indiquez alors pour chaque collection, les contraintes sur la partie droite de la fenêtre. Les contraintes sont les critères que vous imposez pour retenir ou non un texte. **Dominante, Emergente, Marginale, Absente** sont quatre niveaux d'exigence auquel s'ajoute la **simple présence**. Plus le niveau est élevé, plus la collection doit être présente pour que le texte candidat soit retenu. Dans cet exemple, j'ai choisi « Emergente » pour les collections MATERIAUX* et OBJETS-TECHNIQUES* et « Simple présence » pour ACTIVITES ECONOMIQUES*. En cas d'erreur, vous pouvez supprimer une collection sélectionnée en appuyant sur le bouton **Supprimer**.

Une fois votre sélection effectuée, cliquez sur **OK**. Vous revenez alors sur la fenêtre précédente qui vous permet d'**exécuter** le calcul en appuyant sur le bouton destiné à cet effet. Une fenêtre peut alors apparaître. Fermez là et appuyer sur le bouton **Résultats**.

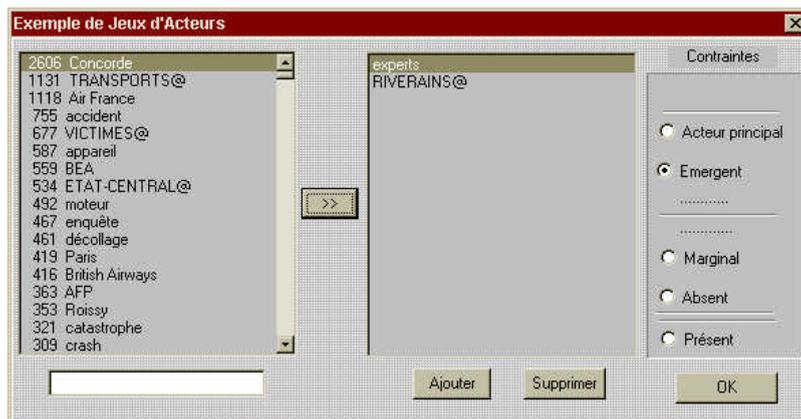
Dans mon corpus, seuls deux textes répondent à ces trois critères. Dans la mesure, où il s'agit de chercher des textes, et non des énoncés, le bouton **Action** ne produit pas d'effet. Il vous faut donc lire le texte. En appuyant sur **Vue/Texte**, vous ouvrez la fenêtre **Mode d'accès – Textes** en étant positionné sur le texte sélectionné. Le bouton **Accès au texte** vous ouvre alors l'éditeur ARIEL, sur lequel nous reviendrons. Pour l'instant, vous pouvez vous contenter de l'utiliser comme un outil permettant d'afficher le document.

En revenant à la fenêtre **Inférences – Répertoires**, vous notez que vous pouvez sauvegarder ou récupérer un ensemble de répertoires respectivement grâce aux boutons **Enregistrer** et **Ouvrir**. Les fichiers ainsi sauvegardés, dont l'extension est REP, peuvent contenir plusieurs répertoires.

2) Les jeux d'acteurs

Le principe est le même que les répertoires. Ici, l'on cherche à repérer des textes en fonction du degré de présence des acteurs et des entités.

Après avoir lancé la commande **Inférences – Jeux d'acteurs**, appuyez sur le bouton **Créer**. En donnant comme nom « Exemple de Jeux d'Acteurs » (ou tout autre nom), vous pouvez définir les éléments le composant grâce au bouton **Editer**. La boîte de dialogue ressemble beaucoup à la précédente :



Les jeux de contraintes fonctionnent de la même façon que pour les répertoires. La notion d'Acteur principal, sur laquelle nous reviendrons lors de la prochaine leçon, est particulièrement exigeante : il faut que l'entité joue un rôle central dans le texte pour que celle-ci soit retenue. Dans cet exemple, j'ai exigé le statut d'acteur principal pour les RIVERAINS@, alors que j'ai simplement choisi « Emergent » pour les experts.

Validez, et procédez comme pour les répertoires : lancez le calcul avec le bouton **Exécuter** et affichez les **Résultats**. Un seul texte dans mon corpus vérifie ces deux critères. Comme précédemment, vous pouvez le faire apparaître avec l'éditeur ARIEL et noter qu'il est assez typique compte tenu des contraintes exigées.

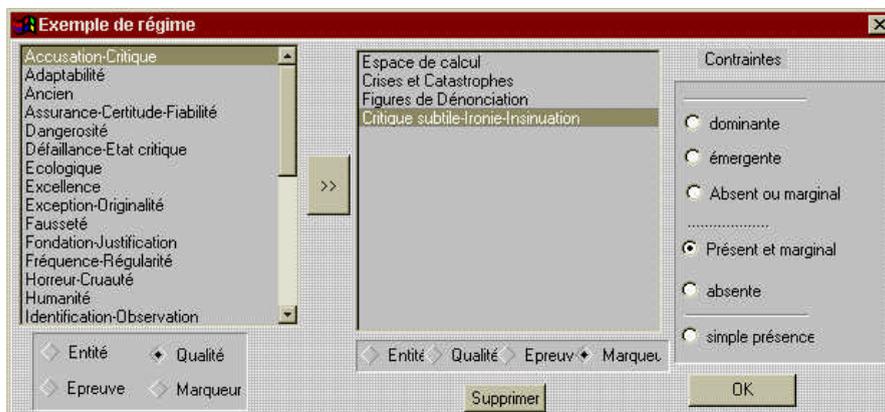
Comme précédemment, on peut enregistrer et récupérer des fichiers contenant plusieurs jeux d'acteurs (leur extension est RAC).

3) Les régimes discursifs

La particularité des catégories par rapport aux autres concepts, selon laquelle on distingue les catégories d'entités, de qualités, de marqueurs et d'épreuves, explique la forme légèrement différente de la boîte de dialogue permettant de construire les régimes discursifs. Mais l'intérêt de cet outil réside précisément dans sa capacité à interroger simultanément les divers types de catégories.

Supposons que l'on cherche à repérer un texte contenant des catégories de différents types à des degrés de présence variables. L'on procède alors comme pour les répertoires et les jeux d'acteurs, en lançant la commande appropriée : **Inférences – Régimes discursifs**.

La démarche est semblable. On crée un régime (choisissons de l'appeler « Exemple de régime »). Puis on l'édite. La fenêtre suivante apparaît :



En bas et à gauche, vous pouvez choisir de faire apparaître les catégories d'entités, de qualités, d'épreuves ou de marqueurs. Vous déplacez la catégorie sélectionnée dans la colonne de droite en indiquant le type de contrainte que vous souhaitez imposer (en bas, pour mémoire, figure le type de catégorie sélectionné). Le choix des différentes contraintes peut être indiqué comme précédemment sur la partie droite. Celles-ci apparaissent dans l'ordre de l'exigence du critère qui conduira à accepter ou réformer les différents textes.

Dans l'exemple présenté, j'ai choisi une catégorie de marqueurs (« critique subtile – Ironie - Insinuation ») et trois catégories d'entités (les autres). « Espace de calcul » et « Figure de dénonciation » sont affectées de la contrainte « simple présence », « Crises et catastrophe » doit être « émergente » alors que « Critique subtile... » est associé à la modalité « Présent et marginal ».

En validant cette boîte de dialogue, vous devez **Exécuter** le calcul et faire apparaître les **Résultats**. Trois textes répondent à ces critères dans mon corpus.

Comme précédemment, vous pouvez les éditer avec ARIEL en appuyant successivement sur les boutons **Vue/Texte** puis **Accès au texte**. Après avoir vérifié qu'ils respectent les contraintes fixées, vous pouvez enregistrer ce régime, avec d'autres que vous pourriez créer, dans un fichier (l'extension est RDI).

4) Les configurations

La vocation de la commande **Inférences – Configurations discursives** est de rechercher des textes selon un ensemble de critères exprimés tant sur des répertoires, des jeux d'acteurs ou des régimes discursifs.

Après avoir lancé cette commande, créez une configuration (on peut l'appeler «exemple de configuration» et demandez son édition. Apparaît une boîte de dialogue comprenant trois boutons principaux qui ouvrent les fenêtres étudiées précédemment : les boutons «jeux de catégories», «jeux d'acteurs» et «jeux de collections» correspondent respectivement aux fenêtres «régimes discursifs», «jeux d'acteurs» et «répertoires».

Cliquons, par exemple sur «jeux de catégories» pour exiger la «simple présence» de deux catégories d'entités («discours écologique» et «catastrophisme») et d'une catégorie de qualité (Accusation-critique). Puis, fixons un «jeux de collections» en imposant la «simple présence» de URBAIN* et PERSONNAGES POLITIQUES*.

La suite est semblable. Validez en cliquant sur OK, lancez le calcul et affichez les résultats. Dans mon corpus, un seul texte répond à l'ensemble de ces critères. Editez-le avec ARIEL pour noter qu'il est vraiment typique d'un discours écologique, porté à la fois sur le pessimisme lié aux catastrophes et sur une critique politique (la question du troisième aéroport), tout en puisant son argumentation sur le vocabulaire urbain (le problème de sa proximité avec les villes est évidemment central).

Il est sans doute utile d'attirer votre attention sur le fait que l'on ne peut choisir qu'un seul critère pour un jeu de concepts donné. Vous comprendrez sur un exemple comment pallier ce problème. Définissons une configuration de la façon suivante :

Jeux de catégories

- Logique financière (catégorie d'entités) : simple présence
- Critique subtile (catégorie de marqueurs) : simple présence
- Marchand (catégorie de qualités) : simple présence

Jeux d'acteurs

- VICTIMES@: simple présence

Jeux de collections

ACTIVITE-ECONOMIQUE* : simple présence

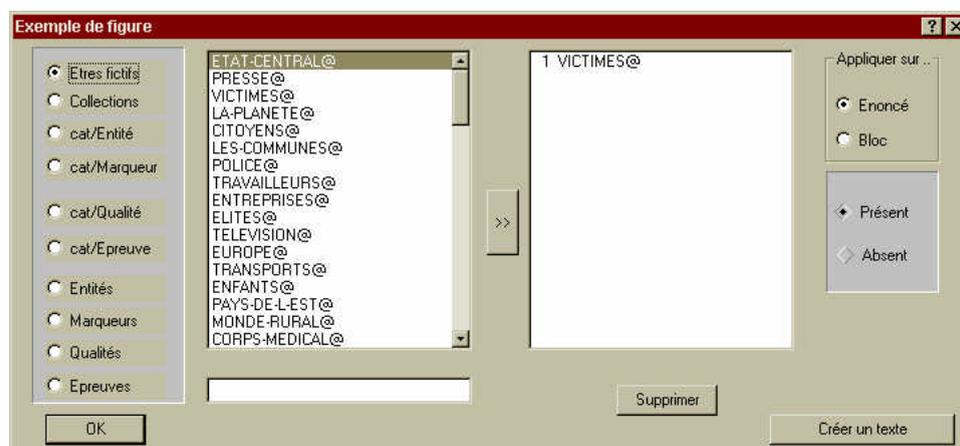
En lançant la requête, vous n'obtiendrez aucun résultat. Pourquoi ? Parce que les jeux d'acteurs et les jeux de collections ne contiennent respectivement qu'un seul critère. Il est facile de contourner ce problème. Il suffit de rajouter dans les deux cas un critère portant sur une très probable absence. Pour les jeux de collections, si vous ajoutez comme critère, l'absence (c'est une modalité des contraintes) d'une collection comme PEINTURE*, JOUEURS D'ECHEC* ou INSTRUMENTS DE MUSIQUE*, vous prenez guère le risque d'écarter des textes tout en indiquant plusieurs critères (il en faut au moins deux). De même, pour les jeux d'acteurs, vous pouvez fixer comme critère l'absence d'une entité n'apparaissant qu'une seule fois dans le corpus (par exemple, « harmonie », « profanation » ou « Papon »). Relancer le calcul pour noter que, cette fois-ci, vous obtenez quelques résultats...

5) Les figures

Les répertoires, les jeux d'acteurs, les régimes discursifs et les configurations offrent donc la possibilité de rechercher des textes répondant à des critères portant respectivement sur les collections, les entités ou être-fictif, les catégories, ou encore l'ensemble des trois.

Les figures, quant à elles, permettent d'identifier, non pas un texte, mais un énoncé (ou un bloc). Du coup, les contraintes sont beaucoup plus fortes, puisque c'est au sein du même énoncé, que l'on doit trouver trace de tous les concepts ou des entités fixés. Afin d'éviter des requêtes ne produisant aucun résultats en raison de critères trop difficiles à satisfaire, l'on ne règle pas les contraintes avec un choix de 5 ou 6 modalités, mais en décidant sur une échelle dichotomique la présence ou l'absence de chaque concept sélectionné.

Prenons un exemple. Lancez la commande **Inférences – Figures** puis créez une figure en définissant un nom (nous choisirons «exemple de figure»), puis éditez-la. La fenêtre suivante apparaît :



Sur la partie gauche, vous choisissez ce sur quoi vous souhaitez voir porter le critère. Après avoir sélectionné les différents éléments, cliquez sur la flèche... et indiquez pour chacun d'entre eux, si vous voulez qu'ils soient « Présent » ou « Absent ».

Dans mon exemple, j'ai cherché un énoncé contenant au moins un représentant de l'être-fictif VICTIMES@, un de la collection PAYS* et un élément de la catégorie d'entités « Dispositifs juridiques ».

En pressant successivement sur les boutons OK, **Exécuter** et **Résultats**, on accède à des énoncés (et non à des textes) en appuyant sur **Action**. Les résultats sont très purs, à l'instar de cet énoncé :

[TEXTE :D:\Concorde\HUM00928.txt / AUTEUR : L'Humanité / date:28/09/2000]

énoncé n° : 4

Mardi, un avocat représentant des familles des victimes allemandes avait indiqué qu ' une plainte contre Continental Airlines était en préparation aux Etats-Unis .

Il vous est possible de générer un texte mettant bout à bout l'ensemble des énoncés répondant aux critères fixés. Pour cela, il suffit d'appuyer sur le bouton **Créer un texte** dans la boîte de dialogue reproduite plus haut. Prospéro vous demande le nom d'un fichier (au format TXT) et une adresse. Vous pourrez, dès lors, ouvrir ce document avec votre traitement de texte.

Enfin, comme précédemment vous pouvez enregistrer et retrouver l'ensemble des figures créées sous un nom de fichier (l'extension est FIG) afin de pouvoir retrouver, d'un corpus à un autre.

Leçon n°4 : J'utilise de nouveaux outils

CE QUE VOUS ALLEZ APPRENDRE :

- *La notion de réseau*
- *La notion d'Acteur Principal*
- *L'accès par les textes et l'éditeur ARIEL*
- *La comparaison de corpus*
- *Le travail sur des interviews*

1. LA NOTION DE RESEAU

Le mot « réseau » suggère la recherche de mots proches d'un élément fixé. C'est sur les entités et les êtres fictifs qu'il est intéressant de travailler car l'idée est surtout de recenser les objets proches.

En fixant une entité de référence, nous allons donc quantifier sa proximité avec les autres entités. Les éléments qui entrent dans le réseau d'une entité ou d'un être-fictif sont tous les éléments (entités ou êtres-fictifs) qui partagent les mêmes énoncés. L'indice d'intensité de la liaison (on préfère parler ici d'intensité ou de force de la liaison plutôt que de poids) est la somme des valeurs prises par la liaison entre l'entité de référence et chaque entité au sein de chaque énoncé.

Tout lien est par défaut représenté par la valeur 1. La valeur 2 est affectée pour marquer la proximité des éléments autour d'une épreuve. Il faut que les entités soient séparées par au moins deux épreuves pour que la valeur soit de 1. Ce modèle permet de structurer un peu plus les réseaux autour des objets liés par des actions, des transformations ou des relations directes et explicites.

Pour comprendre le calcul effectué, le plus simple est sans doute de reprendre notre corpus « Essai » contenant les deux petits textes sur lesquels nous avons travaillé dans la leçon 1. Reproduisons-les :

Texte 1

Ce texte est très heureux d'être le premier à figurer dans votre corpus. Cette antériorité lui procure un sentiment comparable à celui des pionniers débarquant sur une terre inconnue.

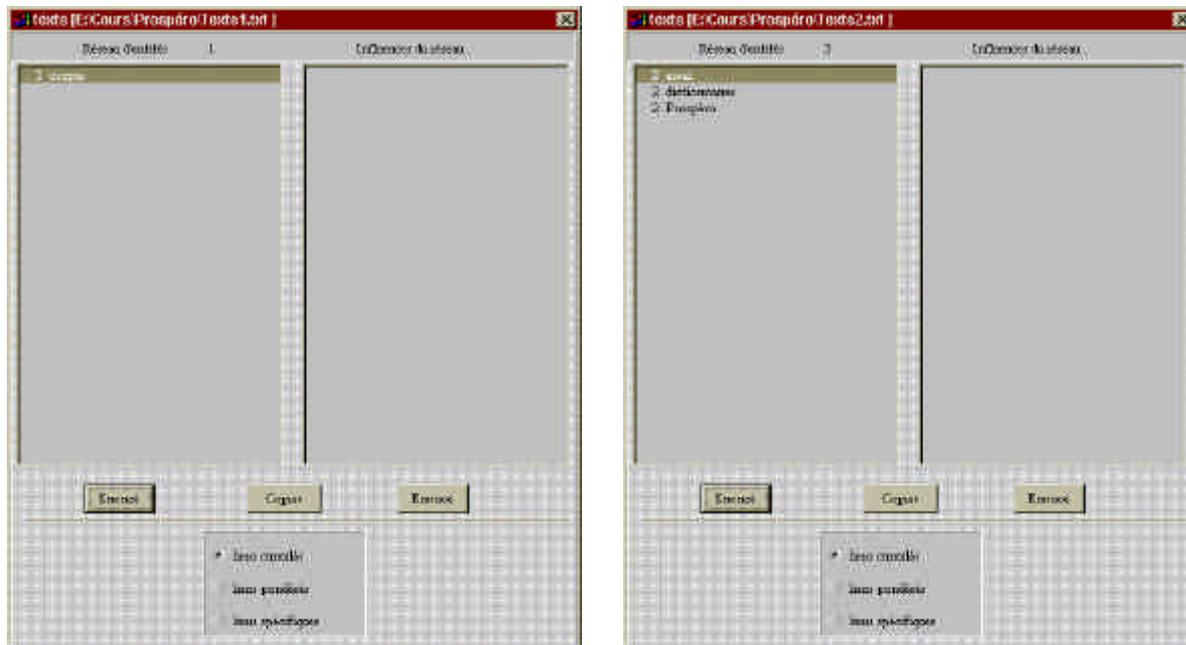
Texte 2

Ce texte est un essai pour comprendre les dictionnaires de Prospéro. Il contient des mots classiques mais aussi des mots n'appartenant pas forcément aux dictionnaires de bases comme aphteuse, pilotis, stéroïde et stertor ou encore des expressions comme « tennis de table ». On peut aussi ajouter un mot n'existant pas encore, comme par exemple, azertyu.

Choisissons par exemple, l'entité « antériorité » qui n'apparaît que dans le texte 1. Les entités liées à « antériorité » sont celles figurant dans l'énoncé, c'est-à-dire, « pionniers », « sentiment » et « terre ». Nous avons ici deux épreuves « procure » et « débarquant ». Seule « terre » est séparée de « antériorité » avec deux épreuves. Son lien est donc de 1 alors que les autres entités reçoivent un score de 2. Vérifions. Après avoir fait apparaître la liste des entités, sélectionnez « antériorité » puis cliquez sur le bouton **Réseaux locaux**. La liste des textes dans lesquelles cette entité est présente se réduit ici au seul TEXTE1. En le sélectionnant et en appuyant sur le bouton **Action**, nous retrouvons les résultats du calcul précédent :



Renouvelons l'expérience avec le mot « texte ». Dans le texte 1, la seule entité figurant dans le même énoncé est le mot « corpus ». Le lien vaut 1 car nous avons deux épreuves entre ces deux entités. Dans le texte 2, « texte » est affecté d'un lien de 2 avec chacune des entités présentes dans le premier énoncé. Vérifions...



Si vous fermez ces fenêtres et que vous appuyez cette fois sur le bouton **Réseau global**, vous noterez que par défaut, Prospero cumule les valeurs de chaque lien.

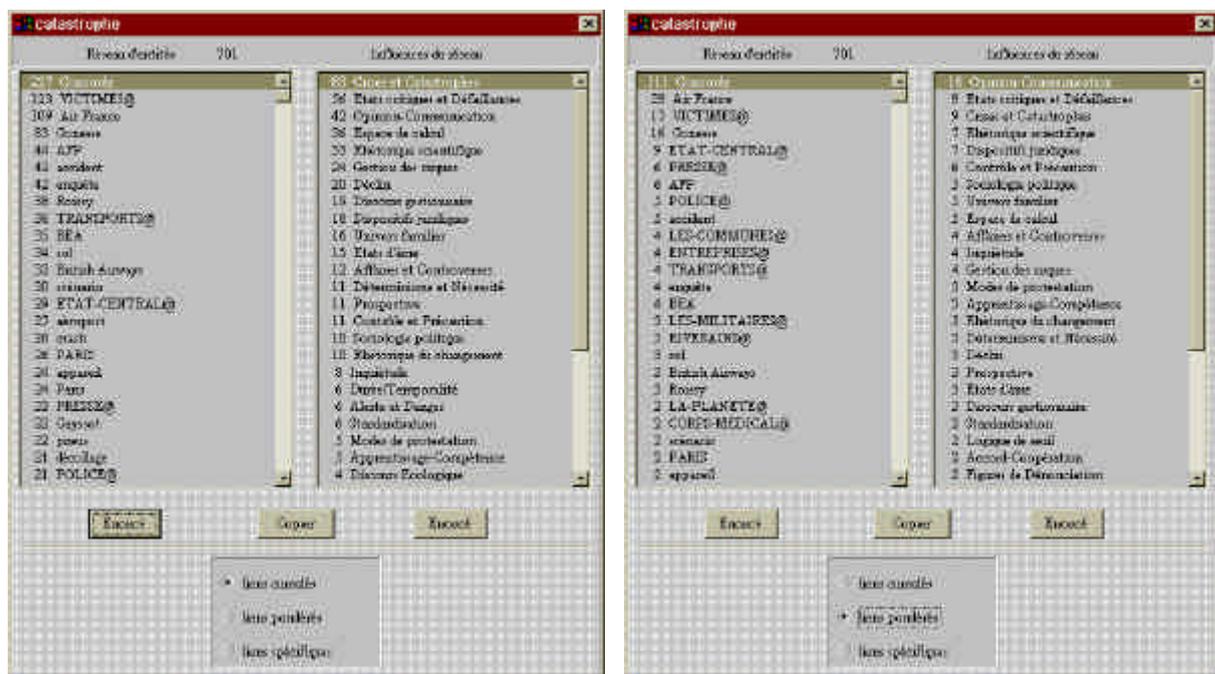


Comme sur les boîtes de dialogues précédentes, vous avez sans doute remarqué que plusieurs fonctionnalités sont disponibles. Bien sûr, le bouton **Énoncé** permet de redescendre dans le

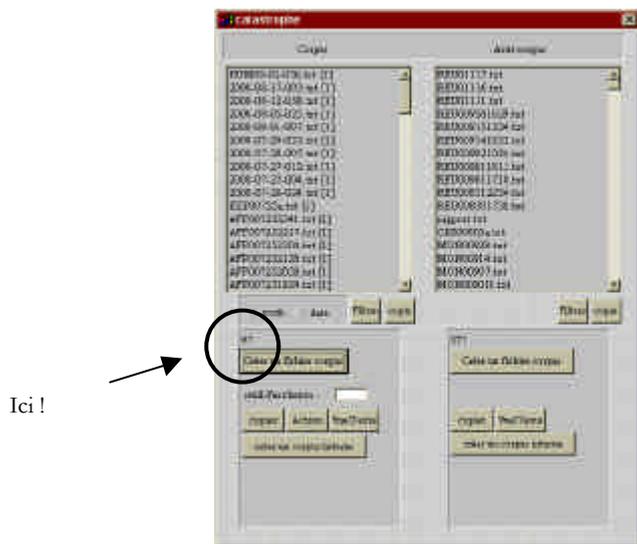
texte : cette possibilité est constitutive d'une position théorique assumée selon laquelle il ne faut jamais perdre prise avec le texte. La partie droite vous livre les catégories dominantes. Il s'agit de la relecture du réseau à travers le filtre des catégories d'entités (le score de chaque catégorie étant le cumul des scores de chacun de ses représentants présents dans le réseau), ce qui donne la propension argumentative du réseau, sa tendance en terme de catégories discursives. Cela permet surtout de sélectionner des textes et des énoncés en se plaçant à un niveau logique supérieur. Le bouton **Enoncé** situé à droite permet d'accéder aux textes contenant l'entité sélectionnée et dans lesquels ces catégories se déploient.

Mais c'est sans doute la partie inférieure de la fenêtre qui mérite d'être explicitée. Vous pouvez en effet choisir d'afficher les liens cumulés (option choisie par défaut), les liens pondérés ou spécifiques. Précisons ces deux derniers modes. Pour obtenir les liens pondérés il faut multiplier le poids résultant du réseau par le pourcentage de textes où les entités considérées (c'est-à-dire celles dont on cherche à évaluer le lien) sont présentes. Illustrons ce calcul par un exemple. Pour avoir des chiffres significatifs, il est préférable de travailler sur un corpus plus gros : ouvrez le projet Concorde.

Si vous cherchez le réseau global d'une entité comme « catastrophe », en demandant successivement les liens cumulés et pondérés, vous obtenez respectivement les valeurs suivantes :



Je me propose d'expliquer pourquoi le lien avec le Concorde est de 111 après pondération. Le calcul est bien sûr le même pour les autres éléments de ces listes. Nous avons 97 textes dans lesquels « catastrophe » et « Concorde » apparaissent dans le même énoncé. Cette information est disponible en appuyant sur le bouton **Enoncé**.

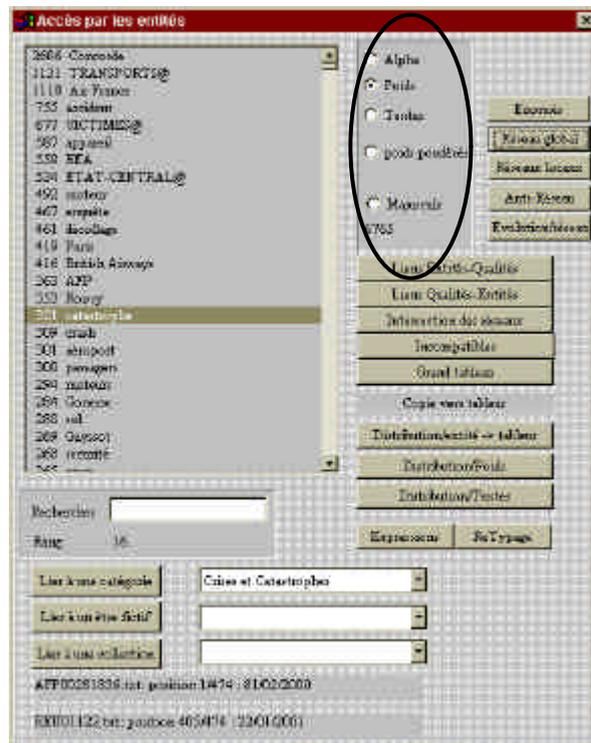


En réalisant la même opération à partir de la fenêtre **Mode d'accès – Entités**, vous noterez que « catastrophe » apparaît dans 198 textes. Il faut donc multiplier la valeur de 227 par le ratio $97/198$, ce qui donne 111.

On perçoit combien cette pondération permet d'apprécier de façon plus « juste » les liens avec les autres entités. Mais l'on peut être ennuyé par un autre problème : des éléments comme « Concorde » ou « Air France » sont systématiquement en tête de liste. Du coup, la lecture des premières entités ne nous apprend rien de très nouveau sur la nature du réseau d'une entité donnée. C'est la raison pour laquelle il est intéressant de regarder ce qui lui est spécifique. Cette opération, qui s'effectue en cliquant sur « liens spécifiques », revient à retirer de la première liste, les entités qui font partie des 100 entités les plus présentes du corpus. Ainsi, on voit apparaître à quoi la « catastrophe » est spécifiquement liée :



Revenons à la fenêtre **Mode d'accès – Entités**. Nous avons jusqu'alors ignoré plusieurs fonctionnalités. Tout d'abord, notons qu'il est possible d'ordonner la liste des entités selon d'autres principes que le nombre d'occurrences.



S'il est sans doute inutile d'expliciter toutes les options de tri disponibles, attardons-nous sur le poids pondéré. Le principe est le même que celui utilisé pour les réseaux pondérés. Il s'agit de

multiplier le score par un ratio calculé en divisant le nombre de textes contenant l'entité sur le nombre de texte composant le corpus.

Nous venons de voir les outils accessibles par les boutons **Réseau global** et **Réseaux locaux**. Attardons-nous à présent sur la notion d'**Anti-réseau**. Il s'agit de l'ensemble des entités et êtres-fictifs qui ne sont jamais dans les mêmes énoncés que l'entité de référence, bien que figurant dans les mêmes textes. Les éléments sont classés par leur poids puisqu'il résume la probabilité a priori d'une connexion avec l'entité source. Par exemple, le fait que Gayssot et les SYNDICATS@ ne soient jamais dans le même énoncé, bien que co-présents dans certains textes, devient une information intéressante. L'anti-réseau revient en somme à structurer les éléments dont la valeur est égale à 0 en terme de calcul des liens.

Il ne faut donc pas confondre l'anti-réseau avec les incompatibles dont la liste est accessible par le bouton correspondant (un peu en dessous). Ici, il s'agit de la liste ordonnée par nombre de textes, des entités ou êtres-fictifs qui ne sont jamais dans les mêmes textes que l'entité source. On les ordonne par nombre de textes pour bien marquer la probabilité a priori que la rencontre ait lieu, ce qui donne, du même coup, une visibilité à l'absence de croisement : par on peut noter que LES-ASSUREURS@ sont incompatibles avec la DGAC (Direction Générale de l'Aviation Civile). Cela marque la non-connexion, ou l'absence de rapprochement de deux dossiers à l'intérieur du dossier...

Il est également possible de calculer l'intersection des réseaux. En cliquant sur le bouton approprié, vous obtenez la fenêtre suivante :



Pour faire apparaître les entités ou êtres-fictifs intervenant dans le calcul (colonne de gauche), il suffit de sélectionner les éléments un à un et d'appuyer sur le bouton **Ajouter**. En cliquant sur

Calculer l'intersection vous obtenez donc les entités figurant dans les deux réseaux ainsi que les catégories communes.

Avant de refermer la fenêtre **Mode d'accès – Entités**, étudions quelques commandes qui peuvent rendre de nombreux services.

Le bouton **Distribution Entités → Tableur** lance la copie de la fréquence de l'entité sélectionnée par rapport à chacun des textes. En collant ce résultat dans votre traitement de texte, ou mieux encore dans votre tableur, vous pourrez demander une représentation graphique de cette distribution.

Celui intitulé **Distribution/Poids** lance une copie de la liste des entités et de leurs poids respectifs. On précise la quantité d'éléments à copier dans la boîte de dialogue suivante. Il convient ensuite de coller cette liste dans le document de votre choix. Notons que cette commande est accessible grâce au bouton droit de la souris lorsque celle-ci est positionnée sur la liste.

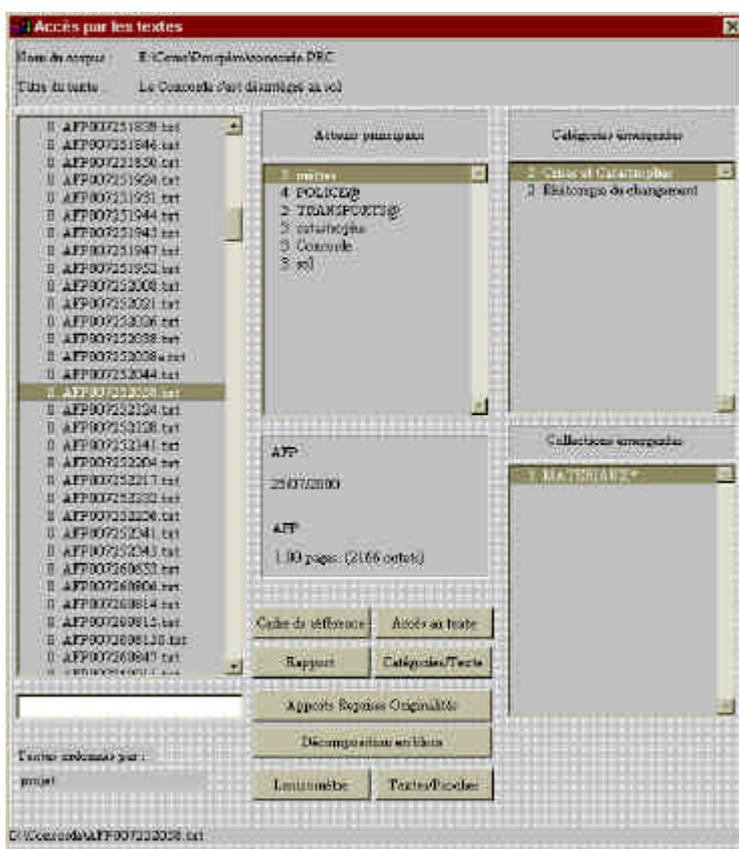
En appuyant sur le bouton **Distribution/Texte**, vous copiez également une partie de cette liste (il faut préciser là encore le nombre d'éléments que vous souhaitez exporter) mais cette fois affectée non pas du nombre d'occurrences mais du nombre des textes dans laquelle les entités apparaissent.

Enfin figurent deux boutons permettant de recenser la liste des entités les plus qualifiées et des qualités les plus souvent attribuées. Il s'agit respectivement des boutons **Liens Entités – Qualités** et **Liens Qualités – Entités**. L'utilisation de ces commandes ne pose pas de problème particulier. Notez que vous pouvez obtenir des résultats semblables avec les formules. Cette fonctionnalité est très utile pour explorer les corpus. En effet, décrire les entités par les qualités qu'elles reçoivent revient à pointer des débats, des controverses et les disputes (par exemple on note que le « développement » est considéré aussi bien comme « prévisible » que « fou », voire de « technologique » ou de « commercial ». Mais il est aussi utile de s'attarder sur les entités peu qualifiées ou celles dont les catégories de qualités sont peu nombreuses : elles indiquent généralement des accords et des consensus. Par exemple, l'entité « aviation » ne reçoit comme qualités que des indications sur la nationalité (« française », « britannique », « américaine »), sur sa nature (« civile ») ou sur des catégories de qualités marquant l'excellence et la fiabilité.

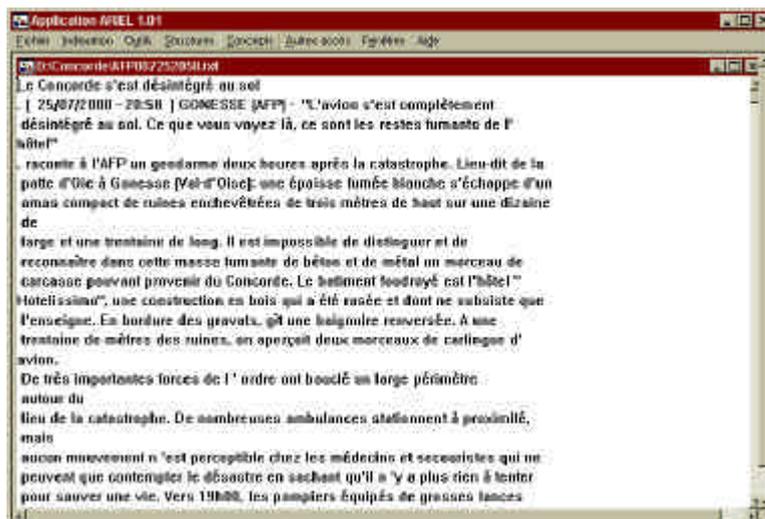
2. LA NOTION D'ACTEUR PRINCIPAL

On appelle Acteur Principal toute entité qui parvient à se hisser en tête de liste d'au moins un texte du corpus de référence. Pour être en tête de liste dans un texte il faut et il suffit que l'entité réalise un des meilleurs scores en termes d'occurrences. Le système s'arrange - ayant incorporé une notion de démocratie représentative assez proche d'une logique floue - pour que les élus soient en nombre suffisant selon la taille du texte et qu'il y ait un écart minimum entre le dernier acteur principal élu et le suivant dans l'ordre décroissant de la liste des entités du texte.

Pour utiliser la notion d'acteur principal, il est conseillé d'ouvrir le corpus «Concorde». En lançant la commande **Mode d'accès – Textes**, vous obtenez une fenêtre que vous avez déjà rencontrée.



Je vous avais invité à utiliser essentiellement deux boutons jusqu' alors : d'une part **Cadre de références** pour saisir les informations dans la base de données, et d'autre part **Accès au texte** pour lancer l'éditeur ARIEL. Ce dernier permet de travailler en partant du texte. Bien qu'il soit sans rapport avec la notion d'acteur principal, attardons-nous sur cet outil.



Lancez la commande **Structures – Mise à jour**. Une boîte de dialogue vous permet de repérer les entités, les éprouves, les marqueurs... du texte, mais aussi les réseaux des différentes entités. Vous disposez aussi de quelques informations sur le pourcentage d'éléments codés et de différents outils permettant de retyper un mot. Fermez cette petite fenêtre et cliquez avec le bouton droit sur n'importe quel mot. Résumez ici les fonctionnalités accessibles :

.....

.....

.....

.....

.....

.....

.....

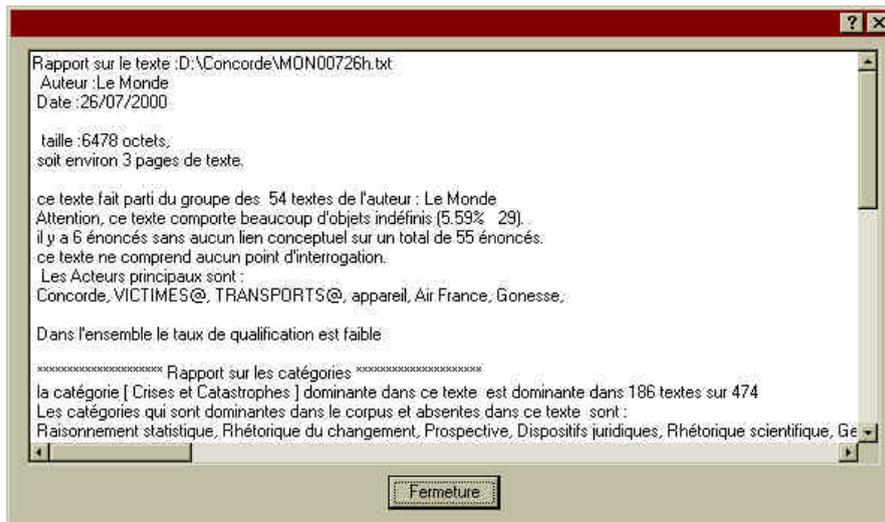
.....

.....

.....

En parcourant les différents menus, vous noterez que vous retrouvez d'autres outils parmi lesquels la création d'expression, la possibilité de lier un élément à un concept, la visualisation des dictionnaires actifs.

ARIEL assure également les fonctions classiques d'un éditeur de texte. Il est notamment possible de rechercher un mot ou une expression (**Outils – Chercher**) et de modifier le texte. Normalement, la déontologie scientifique invite à ne pas changer le matériau empirique sur lequel on travaille (certains pourraient vous faire un procès pour fraude scientifique). Cependant, quand manifestement, vous êtes en présence d'un mot mal reconnu par votre OCR, d'une faute de frappe sans enjeu, ou encore de symboles résultants d'une mauvaise conversion de fichiers, il



Vous l'aurez compris, les fonctionnalités offertes (nous découvrons les autres plus tard) dans la fenêtre **Mode d'accès – Textes** nous aident à étudier le corpus texte par texte. La notion d'acteur principal, dans cette perspective, donne une information pour appréhender rapidement un document.

Mais il est possible de rassembler tous les acteurs principaux pour l'ensemble des textes du corpus. En ce faisant, on peut ainsi avoir une idée assez précise des entités qui jouent un rôle important dans le corpus. Cette opération se réalise à l'aide de la commande **Mode d'accès – Acteurs principaux**.



Cette fenêtre livre énormément d'informations. Sur la partie gauche, figurent l'ensemble des acteurs principaux et le nombre de textes dans lesquels ils apparaissent comme tels. Au centre de cette boîte de dialogue, vous pouvez connaître le premier texte (et sa date) dans lequel cet élément apparaît pour la première fois comme acteur principal. Nous reviendrons sur l'importance des dates lors de la prochaine leçon, mais vous percevez dès maintenant l'intérêt de noter que les SYNDICATS@ se hissent comme acteur principal le jour même du crash. Dans le même texte, lit-on juste en dessous, figurent d'autres nouveaux acteurs principaux : la sécurité, les experts, les pilotes... En dessous de la liste, sur la gauche, vous pouvez afficher la liste des textes dans lesquels SYNDICATS@ est acteur principal (bouton **Liste des textes**) et la configuration émergeant avec cet élément. Le bouton configuration ouvre en effet une fenêtre recensant les entités (et êtres-fictifs), catégories et collections dominantes dans les textes précédents.

En appuyant sur le bouton **Entités concomitantes de la première apparition**, apparaît dans l'espace en haut à droite la liste des entités et êtres-fictifs présents dans le premier texte (au sens chronologique) dans lequel SYNDICATS@ devient acteur principal. Vous pouvez lister les catégories présentes dans ce premier texte en sélectionnant **Cat** (à la place de **Ent/EF** signifiant Entités et êtres-fictifs) et en relançant le calcul grâce au bouton **Entités concomitantes de la première apparition**.

Enfin, la liste des acteurs principaux incompatibles, disponibles en appuyant sur le bouton correspondant, fournit aussi des informations intéressantes. Dire que VICTIMES@, qui figure dans 88 textes comme acteur principal ne l'est jamais avec SYNDICATS@ indique la faiblesse du lien entre ces deux entités. Attention toutefois de ne pas sur-interpréter. Il est faux de dire que l'on ne trouve pas ces deux éléments dans le même texte. En revanche, il n'est pas de texte dans lequel ils seraient l'un et l'autre acteur principal.

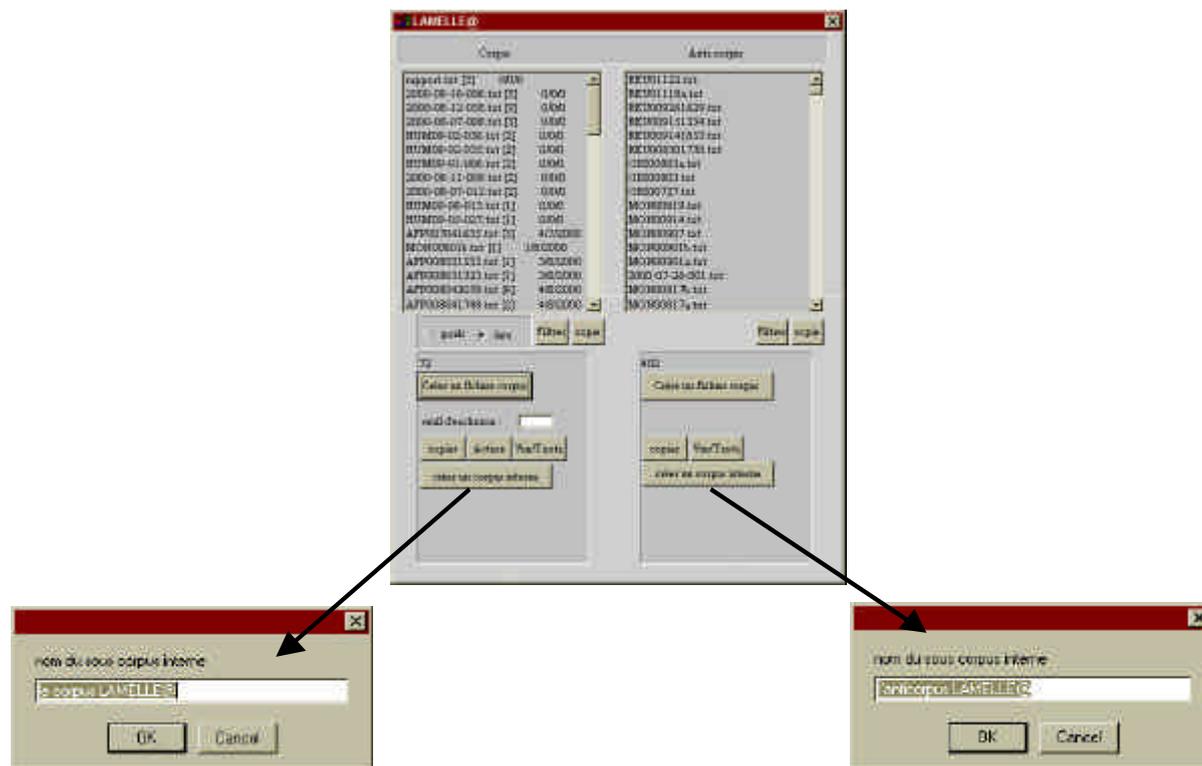
3. COMPARER

Il est souvent utile de comparer un groupe de textes à un autre du point de vue des entités, des qualités,... mais aussi des catégories, des collections, des personnes. Les critères permettant d'isoler un ensemble de texte peuvent être variés : il peut s'agir de la présence d'un mot, d'une expression, d'une catégorie, d'une collection,... Dans tous les cas, l'on constitue des sous-corpus.

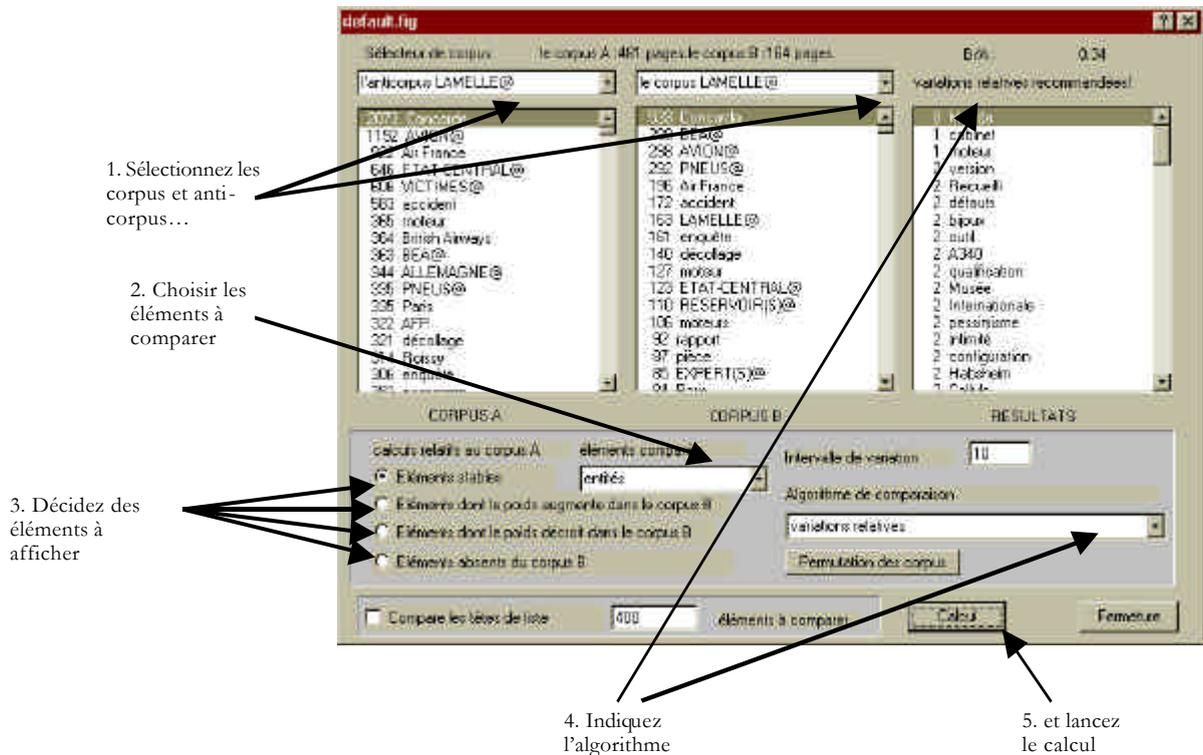
Pour comprendre cette démarche, très courante, prenons un exemple. Admettons que nous cherchons à apprécier l'importance de l'apparition de la lamelle métallique, qui rappelons-le, serait responsable de l'éclatement du pneu et donc de l'accident. Pour cela, on peut diviser le corpus en deux sous corpus : l'un rassemblera tous les textes dans lesquels un représentant de l'être-fictif LAMELLE@ est présent et l'autre, tous les autres.

Pour ce faire, sélectionnez LAMELLE@ dans la fenêtre **Mode d'accès – Entités**, puis cliquez sur **Enoncés**. Apparaît une boîte de dialogue que vous connaissez bien. Cliquez sur les deux boutons **Créer un sous-corpus interne**, l'un étant situé sous la colonne « Corpus » (à gauche), l'autre sous la colonne « sous-corpus » (à droite). Prospéro vous propose de nommer ces deux sous-corpus, respectivement : « le corpus LAMELLE@ » et « l'anti-corpus LAMELLE@ ». Acceptez ce choix.

Notez que vous pouvez également constituer un « fichier corpus ». Dans ce cas, il vous faudrait donner un nom de fichier. L'extension PRC indique qu'il s'agit bien d'un nouveau projet. Rien ne vous empêche alors de ne travailler que sur une partie de vos textes, éventuellement pour comparer avec une autre. Un sous-corpus interne, en revanche, n'existe pas sur le disque. L'on ne peut donc le sauvegarder. Cet inconvénient constitue aussi sa force puisqu'il se crée très rapidement. Travaillons donc avec des sous-corpus internes.



Fermez ces fenêtres pour lancer la commande **Inférences – Opérations sur des sous-corpus – Comparaisons**.



1. Choisissez de quel sous-corpus sera le A (et donc l'autre, le B) ! mettre par exemple, l'anti-corpus LAMELLE@ à gauche, et le corpus à droite) grâce aux menus déroulants

2. Sélectionnez les éléments que vous souhaitez comparer. Vous avez le choix entre une comparaison des entités, des catégories d'entités, des catégories de qualités, des catégories de marqueurs, des catégories d'épreuves, les acteurs principaux, les collections et les marqueurs. Pour faire apparaître les derniers, appuyez sur les petites flèches :



Choisissons, par exemple, les entités. Apparaît alors sur les deux premières colonnes, les entités dans l'anti-corpus et dans le corpus (si en 1, vous avez fait apparaître les sous-corpus dans cet ordre).

3. Vous disposez de plusieurs choix. Si vous sélectionnez « Eléments stables », Prospéro n'affichera que les entités qui ne subissent pas de grande variation. Vous pouvez indiquer la variation que vous jugez significative en modifiant la valeur de la variable : « Intervalle de variation ». Généralement, on conserve 10 (il s'agit d'un pourcentage).

4. Il vous faut choisir un algorithme de comparaison. Deux options s'offrent à vous : variations relatives ou absolues. Le choix est généralement guidé par le ratio entre la taille des deux sous-corpus. Dans notre cas, l'anti-corpus LAMELLE@ est équivalent à 481 pages, alors que le

corpus LAMELLE@ se déploie sur 164 pages, ce qui revient à un rapport de 0.34 (cette information apparaît en haut à droite de la fenêtre). Il paraît donc logique de tenir compte de cette différence de volume, ce sans quoi, les scores de la plupart des entités baisseraient simplement en raison d'une diminution de la taille du sous-corpus. L'algorithme « variations relatives » calcule les effectifs attendus (en cas de volume semblable) pour évaluer des écarts : il est détaillé en annexe.

5. Le bouton **Calcul** lance l'opération : les résultats s'affichent dans la colonne de droite. En double cliquant sur un élément de cette liste (que vous pouvez copier grâce au bouton droit de la souris), vous faites apparaître les scores de cette entité dans le corpus et le sous-corpus.

La logique est la même selon que vous souhaitez comparer des catégories, des acteurs principaux, des collections...

Concernant les éléments à afficher, vous disposez de plusieurs options :

- la liste des entités relativement stables (5% ou 10% selon l'exigence de l'utilisateur) ;
- la liste des entités dont le poids augmente quand on passe de A à B : les entités sont ordonnées selon l'importance de leur variation. Notez qu'une entité peut augmenter de plus de 100% du poids obtenu dans le corpus de référence, par exemple une entité qui passerait d'un poids de 10 à un poids de 100... ;
- la liste des entités dont le poids diminue quand on passe de A à B : les entités sont ordonnées de même selon l'importance de leur variation (contrairement au cas précédent une entité ne peut pas diminuer de plus de 100% puisque cela correspond à sa disparition pure et simple) ;
- la liste des entités qui disparaissent quand on passe de A à B. Dans ce cas le poids observé dans B est égal à zéro.

Pour des raisons de symétrie, on a créé une fonction de permutation qui permet d'opérer la comparaison dans le sens inverse, c'est-à-dire en prenant le corpus B

L'interface mise en place dans le logiciel permet en outre de filtrer la liste des éléments à comparer en travaillant par exemple sur les premiers éléments des listes. En bas à gauche de la fenêtre, vous pouvez indiquer que le calcul ne s'opèrera que sur les 400 entités (si vous travaillez sur cet élément). D'une part, cela accélère le calcul sans perdre pour autant des informations importantes. D'autre part, il convient de se méfier d'une variation sur un élément marginal qui en passant par exemple de 1 à 2 occurrences, sera traduite par une augmentation de 100%.

Dans l'exemple choisi ici, on notera que l'apparition de la LAMELLE@ ne change rien sur certains éléments. En particulier, les débats sur le certificat de navigabilité (l'être-fictif CERTIFICATS@ regroupe plusieurs représentants) sont aussi nombreux (relativement). En revanche, il est beaucoup plus question de RESERVOIR(S)@, de PNEUS@ (le poids augmente). Parmi les éléments diminuant, figurent les RIVERAINS@, les ASSOCIATIONS@, le PARLEMENT@... Disparaissent complètement du corpus, l'émotion, l'église,... c'est-à-dire les éléments proche de l'accident (qui a eu lieu 10 jours auparavant) et ceux liés à une autre controverse, notamment celle autour du TROISIEME-AEROPORT@,...

Cet exemple montre comment à partir des éléments de base mis en place dans le logiciel on peut proposer de «tendre l'épreuve» pour aller vers une sorte de métrologie textuelle dotée d'une forte capacité d'objectivation.

4. LES INTERVIEWS

Le chercheur en sciences sociales travaille souvent par entretien. Prospéro peut bien sûr traiter ce matériau. Les fonctionnalités vues jusqu'alors permettent d'analyser finement les textes mais il faut prendre quelques précautions pour ne pas voir les listes polluées par des éléments propres à la retranscription.

Notamment, en faisant précéder chaque question de votre nom (ou d'un tiret suivi du mot « enquêteur » ou « question ») et chaque réponse du nom de la personne interrogée, il est probable que ces éléments figurent en haut de la liste des entités.

Vous pouvez certes vous affranchir de ce type de problème en remplaçant les « Enquêteur » ou « Réponse » par les initiales des individus. Il vous suffira alors de les typer comme des mots-outils. Mais l'on peut aussi tenter de garder cette information pour savoir qui parle. Dans ce cas, vous pourrez utiliser le module interview et bénéficier de quelques fonctionnalités intéressantes. Elles sont particulièrement utiles lorsque votre texte contient plusieurs locuteurs qui parlent à tour de rôle ou au contraire en se coupant la parole...

Prenons un exemple. Je vous propose de taper ce petit texte, qui constituera le seul élément d'un nouveau projet.

(LOC ENQUETEUR) : Comment avez-vous commencé la sociologie ?

(LOC ACTEUR1) : C'est ma mère qui m'y a incité.

(LOC ENQUETEUR) : Parlez-moi de votre mère.

(LOC ACTEUR1) : Elle adore M. Machin et ses chroniques à la radio. Je ne connais d'ailleurs pas exactement la nature de leurs relations.

(LOC ENQUETEUR) : Quand en avez-vous pris conscience ?

(LOC ACTEUR1) : Très tardivement. Je crois que j'avais 15 ans.

(LOC ACTEUR2) : Je m'insurge contre le terrorisme exercé M. Machin !

(LOC ENQUETEUR) : Ce n'est pas le problème

(LOC ACTEUR2) : Si précisément.

(LOC ACTEUR1) : Pour moi, ce n'est pas très important.

(LOC ENQUETEUR) : Parlez-moi de l'influence de M. Machin sur vos travaux...

Pour identifier le locuteur, il faut respecter la syntaxe suivante : vous devez taper LOC (en majuscule) suivi du nom de la personne (ou de ses initiales), le tout doit être placé entre parenthèses et suivi de deux points (:)

Si vous avez déjà retranscrit l'intégralité de vos entretiens, il vous suffit avec votre traitement de texte de lancer la commande **Edition – Remplacer** pour substituer tous vos « Enquêteur » par « (LOC ENQ) : »...

Ce travail étant réalisé, choisissez **Module Interview** dans le menu **Mode d'accès**.

En appuyant sur le bouton **Repérage des locuteurs dans le corpus**, les différentes colonnes de la fenêtre se remplissent.



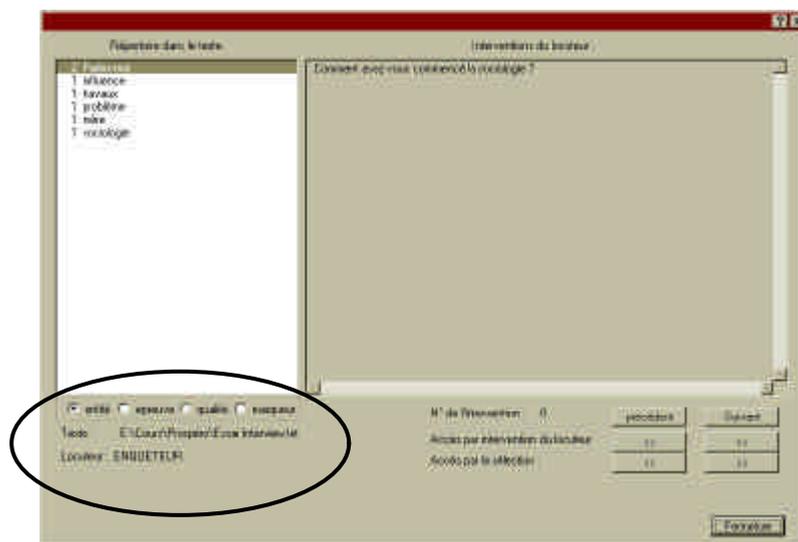
- Sur celle de gauche, figurent les locuteurs identifiés dans le corpus. Vous pouvez sélectionner celui dont vous souhaitez étudier le discours.

- La colonne suivante affiche les entités, les épreuves, les qualités ou les marqueurs (il faut indiquer son choix en dessous de cette colonne) utilisés par le locuteur sélectionné.
- La troisième fait apparaître les textes dans lesquels apparaît le locuteur sélectionné et son nombre d'interventions. Ici, seule cette valeur varie puisque nous n'avons qu'un seul texte.
- Les résultats affichés dans la colonne de droite dépendent, du locuteur sélectionné et de l'option de visualisation choisie. Si vous souhaitez en changer, il vous faudra appuyer à nouveau sur le bouton **Repérage des locuteurs dans le corpus**. Cet outil permet de rendre compte de la succession des prises de paroles.

Chacune des listes figurant sur ces quatre colonnes peut être copiée vers votre traitement de texte ou votre tableur grâce au menu contextuel (accessible par le bouton droit de la souris). Il suffit de choisir **Copier liste** et de préciser le nombre d'éléments à copier.

Enfin, vous disposez d'un autre bouton : **Visualisation des interventions par texte**. Les résultats affichés dans cette nouvelle fenêtre dépendent du texte choisi (ici, nous n'en avons qu'un) et du locuteur sélectionné.

Sur la partie gauche, vous pouvez faire apparaître les entités, les qualités, les épreuves et les marqueurs utilisés par le locuteur sélectionné pour un texte donné (la boîte de dialogue précédente affichait ces informations mais sur l'ensemble du corpus). Ces options sont disponibles en bas à gauche de la fenêtre.



Sur la partie droite, vous pouvez dérouler le texte intervention par intervention grâce aux boutons **Précédent** et **Suivant**. Les flèches situées juste en dessous (<< et >>) permettent de voir défiler les interventions du seul locuteur.

5. EXERCICES

Exercice 1 :

Identifier le réseau spécifique de Gayssot

Exercice 2 :

Sur le corpus Concorde, faire la liste des entités ne figurant pas dans les mêmes textes que ceux contenant l'être-fictif VICTIMES@

Exercice 3 :

Sur le corpus Concorde, faire la liste des entités ne figurant pas dans les mêmes énoncés que ceux contenant la DGAC

Exercice 4 :

Sur le corpus Concorde, lister les catégories dominantes des textes contenant la DGAC

Exercice 5 :

Sur le corpus Concorde, dresser la liste des textes dans lesquels les ASSUREURS@ sont acteur principal. Identifier les collections et les catégories d'entités dominantes dans cette liste de texte.

Exercice 6 :

Quelles sont les catégories d'entités dominantes lors de l'apparition de « sécurité » comme acteur principal.

Exercice 7 :

Dresser la liste des 10 premières catégories d'entités déployées dans le texte RAPPORT.TXT sur corpus Concorde.

Exercice 8 :

Sur le corpus Concorde, créer deux sous-corpus : l'un contenant la catégorie d'entité « Gestion des risques », l'autre ne la contenant pas.

Comparer ces deux sous-corpus du point de vue des entités et des personnes.

CE QUE VOUS ALLEZ APPRENDRE :

- *Repérer le déploiement temporel d'une notion et le temps du corpus*
- *Identifier les dates dans un corpus*
- *Recenser les éléments précurseurs et comparer entre un « avant » et un « après »*
- *Etudier l'évolution d'un réseau*
- *Découper en périodes*

1. LE DEPLOIEMENT TEMPOREL DU CORPUS

Prospéro accorde au temps une place centrale. Les fonctionnalités liées à la temporalité résultent d'une position épistémologique invitant à prendre au sérieux le temps que se donnent les acteurs. On a vu, les leçons précédentes, comment le logiciel permettait de repérer des mots ou expressions comme « désormais », « de plus en plus », « à un moment donné », de recenser les contextes sociaux dans lesquels ils sont utilisés, et donc, d'identifier les moments de rupture dans leur narration des événements passés, les doutes, les façons d'envisager le futur...

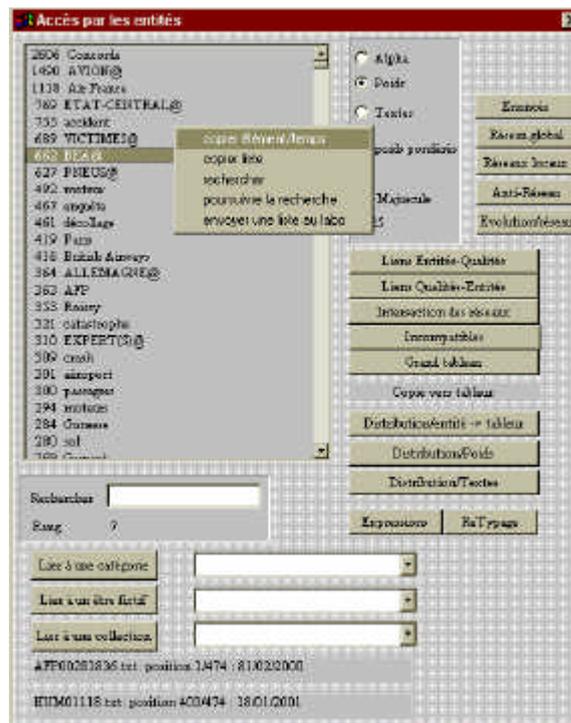
En dehors de ces outils, Prospéro offre des fonctionnalités propres à traiter la dimension temporelle du corpus. La seule contrainte est bien sûr d'avoir complété le champ « date » dans les références externes (cela va de soi, mais cet oubli conduit parfois à une incompréhension étonnante).

Tout d'abord, il est possible de regarder comment se distribuent vos textes selon les dates. Pour cela, il vous suffit de lancer la commande **Temporalités – Temps du corpus**. Vous pouvez alors choisir l'unité : prenons ici une période mensuelle.



Un message vous confirme l'opération. Ouvrez votre traitement de texte ou votre tableur puis faites **Edition – Coller**. Avec un tableur, il est ensuite très facile de réaliser un histogramme ou toute autre représentation graphique¹.

Il est possible de faire le même travail sur n'importe quel élément du corpus. Dans la fenêtre **Mode d'accès – Entités**, sélectionnez BEA@. Puis cliquez avec le bouton droit de la souris pour choisir **Copier Élément/temps**.



Vous retrouvez la boîte de dialogue précédente. Choisissez, par exemple : 1 jour. La démarche pour coller dans une application de votre suite bureautique est semblable.

¹ En cas de difficultés, vous pouvez vous reporter à TRABAL P., *Pratiquer le tableur...*, Vuibert, Paris, 2000.

Un autre outil est très pratique. Il permet de recenser toutes les dates du corpus. Il ne s'agit pas ici des dates auxquelles ont été écrits les textes mais des dates mentionnées dans les documents. Ces informations sont précieuses car elles permettent de repérer les moments importants pour les acteurs sociaux. Lorsqu'il s'agit de dates passées, on peut recenser les appuis historiques, les précédents, les événements considérés comme marquants... Mais il est des corpus où l'on engage fortement l'avenir : nos descriptions méritent alors d'identifier l'horizon temporel des acteurs et leur façon d'envisager le futur.

Lancez la commande **Temporalités – Datations dans les textes**. Vous avez la possibilité entre quatre options : Prospéro peut chercher

- les dates s'exprimant sous forme de jour : l'on récupère alors des nombres allant de 1 à 31 présentant la particularité d'être suivi par un mois
- les dates sous forme de mois : s'il affiche les nombres entre de 1 et 12, il prend en compte des expressions comme « en octobre 2000 », ou encore « 9/00 ».
- les dates en années : il recense tant les formes du type 1998 que celles prenant en compte les deux derniers chiffres (98).
- toutes les dates : il prend alors en considération toutes les formes précédentes

Du coup, vous trouverez inévitablement un peu de bruit. Une expression comme « entre 1800 et 2000 dollars » risque de le laisser supposer que 1800-2000 constitue une plage temporelle. Mais s'il faut rester vigilant, cet outil permet cependant d'identifier l'horizon du corpus. Parmi les dates éloignées dans le futur, on a 2025, 2050 (vous pouvez vérifier qu'il s'agit bien d'échéances prévisionnelles...). Comment ? On clique sur le bouton **Énoncé** situé en bas à gauche. La colonne de droite, vous l'aurez compris, permet de recenser les principales entités dans les énoncés contenant la date sélectionnée.

Enfin, vous noterez que vous pouvez classer la liste des dates, soit par ordre des dates (pour voir le temps que se donne les acteurs), soit par poids (pour repérer les dates marquantes).

Mais vous pouvez également étudier l'impact d'un événement sur le dossier. Le 31 août 2000, le BEA rend son premier rapport sur l'accident. On peut essayer d'évaluer l'impact d'un tel fait sur le dossier. Pour cela, il faudrait séparer un « avant » et un « après ». Un outil réalise cette tâche : lancez la commande **Temporalités – Périodisation manuelle**. Compléter la date du rapport :



Pour séparer notre corpus en deux, une date suffit. Mais rien ne vous empêche d'identifier plus de deux périodes ; dans ce cas, il vous faudra compléter le début de période. Cliquez sur **Création d'un corpus**.

Une fenêtre que vous connaissez bien apparaît. Il suffit de créer les sous-corpus et de lancer le module de comparaison **Inférences – Opérations sur les sous-corpus – Comparaison** (voir leçon n°4).

2. L'EVOLUTION D'UN RESEAU

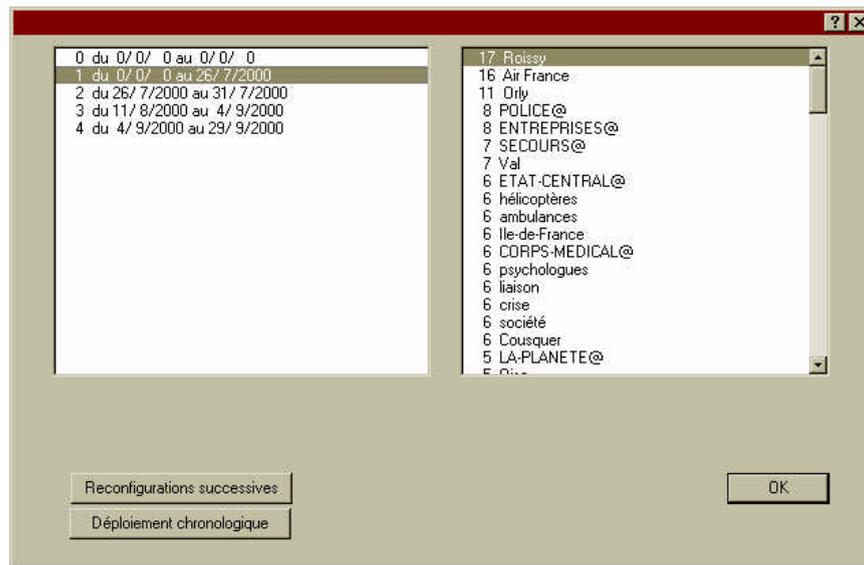
Lors de la leçon précédente, nous avons étudié la notion de réseau. Pour résumer, rappelons qu'il s'agit de l'ensemble des entités proches d'une entité donnée. Les liens se mesurent à l'intérieur de chaque texte, mais peuvent être également cumulés.

Souvent, nos acteurs bougent. Aussi, il semblait intéressant de rendre compte de leur évolution dans le temps. Il est clair que cette perspective est surtout intéressante sur des corpus de longue durée dans lesquels la dimension transformationnelle ou historique est centrale. C'est moins naturellement le cas sur des corpus d'interviews ou sur des corpus se développant sur une période très courte.

Pour rendre compte de l'évolution de nos acteurs, Prospéro ordonne les textes par ordre chronologique, puis calcule les évolutions des têtes de réseau de l'entité source pour chaque texte en considérant qu'il y a reconfiguration du réseau dès lors que plus de 75% des têtes de liste ont changé. Il calcule du même coup la période pendant laquelle le réseau est relativement stable. Ainsi, l'on dispose d'une périodisation non pas réalisée comme précédemment à l'initiative du chercheur (pour tester l'une de ses hypothèses) mais à partir des éléments proches de l'entité dont on veut comprendre l'évolution.

Prenons l'exemple de l'institution « Aéroport de Paris » ou ADP. Il est souhaitable, si ce n'est pas encore fait, non seulement de typer l'expression, mais aussi de construire un être-fictif (ADP@) pour rassembler toutes les manières de nommer cette société.

Après avoir fait apparaître la liste des entités (**Mode d'accès – Entités**), sélectionnez cet être-fictif. Cliquez alors sur le bouton **Evolution/réseau**, puis sur **Reconfigurations successives**.



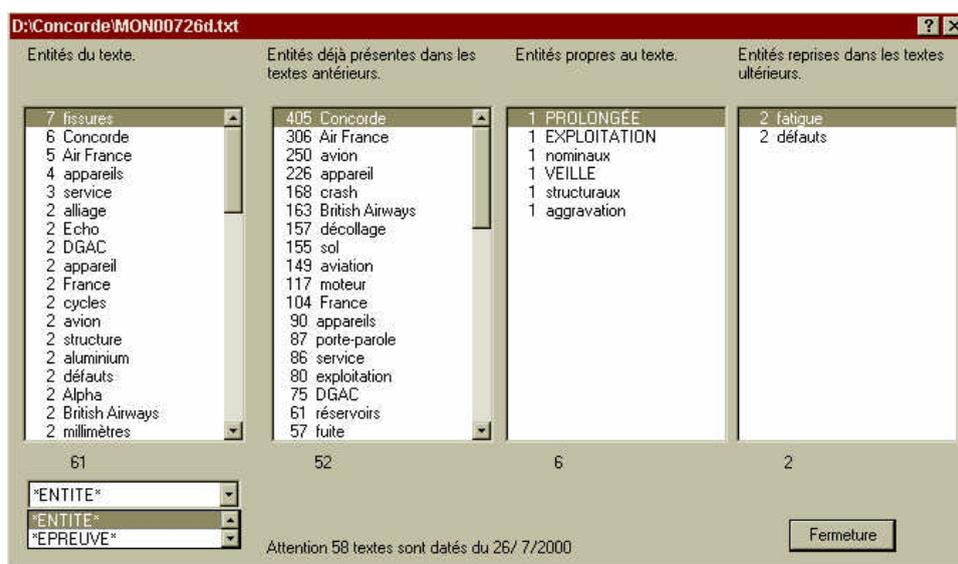
Prospéro identifie 4 phases en raison du fait que certaines dates n'ont pas été complétées dans mon corpus. Le vôtre est sans doute plus complet et mieux organisé. Dans tous les cas, vous pouvez identifier une période initiale allant de l'accident jusqu'au lendemain. La colonne de droite affiche le réseau des ADP@ sur ces deux jours : il est question du lieu, des SECOURS@, de la POLICE@, du CORPS-MEDICAL@ et autres « psychologues ». Si vous sélectionnez à présent, la période du 26/7 au 31/7, vous noterez que la tête du réseau fait place à de nouvelles entités comme la « Commission », « l'administrateur », la « gestion », les « mesures »... Puis en cliquant sur la période suivante, vous noterez que la tête du réseau change à nouveau : il est question de la fameuse LAMELLE@. On évoque la responsabilité des ADP@ dans cet accident, ce qui explique la présence du BEA@, d'une « inspection », du « rapport »... Puis dans la dernière période (en tout cas pour mon corpus), on voit apparaître la question du PNEU@, de la responsabilité de « Goodyear » (le fabricant du pneumatique ayant éclaté), de leur responsabilité...

3. LE TEXTE PRECURSEUR

Il est une fonctionnalité que nous n'avons pas abordée lorsque nous travaillions sur la fenêtre **Mode d'accès – Textes**. Nous avons en effet laissé un outil permettant de savoir ce qu'un texte apporte dans le corpus. Cela renvoie bien sûr à la temporalité puisque cette opération revient à prendre la liste des entités (respectivement des qualités, des épreuves ou des marqueurs d'un texte donné et de la comparer avec celle des entités (respectivement des qualités, ...) des textes antérieurs et postérieurs.

Dans la fenêtre **Mode d'accès – Textes**, choisissez un texte, si possible ni le premier, ni le dernier. Prenons, par exemple MON00726d. Appuyez sur le bouton **Apports Reprises Originalités**.

Dans la boîte de dialogue qui s'ouvre, choisissez à l'aide du menu déroulant en bas à gauche, de travailler sur les entités (par exemple).



Sur la colonne de gauche, on peut alors lire les entités figurant dans le texte sélectionné, ainsi que leur nombre. Les colonnes suivantes affichent respectivement les entités présentes antérieurement (il faut rester prudent lorsque de nombreux textes sont datés du même jour), celles propres au texte sélectionné, et celles qui seront reprises ultérieurement. Ici, on note que c'est à partir de ce texte que l'on évoque, par exemple, la « fatigue ». C'est un élément qui peut être intéressant. Il peut être aussi pertinent de noter que certains textes n'apportent rien (en terme d'entités, de qualités,...) au corpus. C'est souvent le cas des textes de synthèse, parfois de

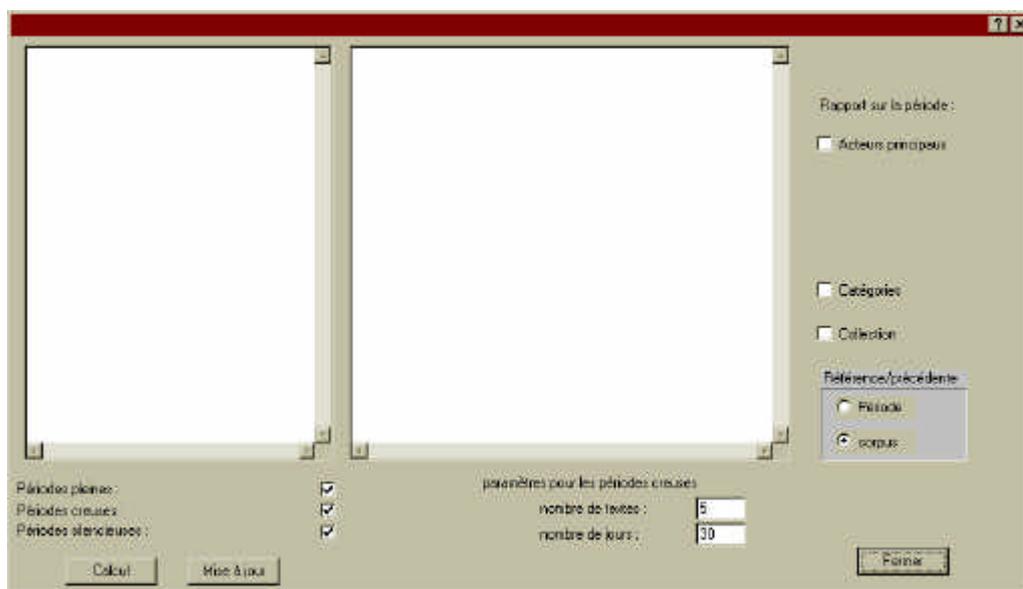
discours politique. Mais un texte prétendant résumer un dossier qui rajoute de nombreux éléments non disponibles précédemment, mérite aussi une attention particulière.

Notez que le bouton droit de votre souris vous offre de nombreuses possibilités, parmi lesquels l'exportation via un copier-coller vers votre traitement de texte ou votre tableur, la recherche, le tri...

4. LES PERIODES ET LES EPISODES

Vous connaissez deux façons d'obtenir une segmentation de votre corpus en période. L'une est manuelle : vous décidez des dates pertinentes et lancez la commande **Temporalités – Périodisation manuelle**. L'autre est réalisé par Prospéro et résulte d'un travail sur les têtes de réseau d'une entité donnée.

Il existe une autre façon d'obtenir une périodisation. Lancez la commande **Temporalités – Calcul des périodes**.



Sur la partie droite de cette fenêtre, vous indiquez votre souhait de travailler sur les acteurs principaux, sur les catégories ou les collections. Par exemple, cochez la case « Collection ».

Il faut distinguer à présent les périodes pleines, creuses et silencieuses. La période est dite silencieuse lorsque aucun texte ne figure dans cet intervalle de temps. L'on distingue les périodes creuses des pleines en fixant le nombre de textes et le nombre de jours dans la zone « paramètres

pour les périodes creuses ». Décidons par exemple qu'une période est creuse lorsque l'on a moins de deux textes par jour. Le bouton **Mise à jour** lance le calcul permettant d'identifier le nombre périodes pleines, creuses et silencieuses, soit dans mon corpus, respectivement 22, 41 et 15. Après avoir sélectionné simplement « périodes pleines » pour recenser les catégories présentes à chacune des périodes, il est possible de cliquer sur n'importe quelle période pour voir afficher, sur la partie droite de la fenêtre les collections présentes, celles qui ont fait leur apparition depuis la précédente, celles qui ont disparu.

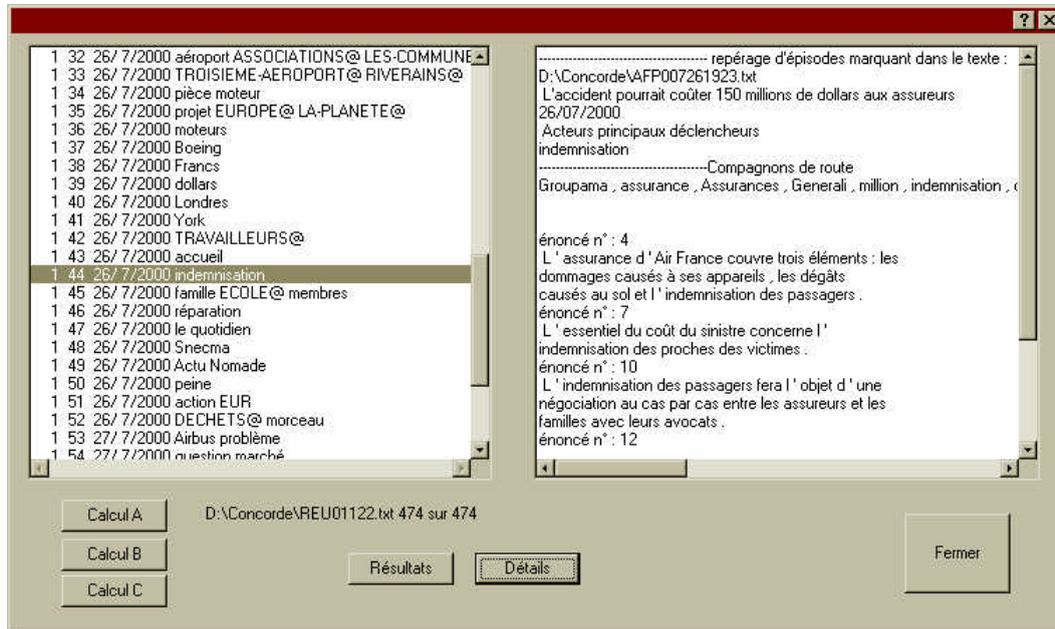


Ces renseignements (figurant sur la colonne de droite) peuvent être sélectionnés avec la souris et copiés : vous pouvez donc les récupérer dans votre traitement de texte. En cliquant avec le bouton droit sur la liste, vous noterez qu'il est également possible d'exporter les différentes périodes sur votre tableur (et donc de réaliser par exemple un histogramme). Enfin, de la même façon, on peut accéder aux textes constituant cette période. Du coup, il devient facile de créer des sous corpus, de les comparer...

Le rapport sur la période peut s'effectuer sur les catégories et sur les acteurs principaux. Cet outil est donc très utile pour l'historien.

Vous disposez enfin d'une dernière commande en lien avec la temporalité. Il s'agit d'un algorithme capable d'identifier les événements marquants grâce aux épisodes. Lancez la commande **Temporalités – Episodes**.

En cliquant sur **Calcul A**, Prospéro démarre un calcul visant à distribuer les groupes d'entités, relativement proches les unes des autres, en fonction des dates. (A VERIFIER). En cliquant sur un des « épisodes », nous obtenons sur la partie de droite, les énoncés significatifs dans lesquels ces entités apparaissent.



A REVOIR

