

# Typologie des fonctionnalités textométriques selon un point de vue utilisateur : illustration par leurs implémentations dans des logiciels

Bénédicte Pincemin (*éditeur*)

Les différents collaborateurs ayant participé aux descriptions sont mentionnés au niveau de chaque logiciel.

*Cette note technique a été initiée dans le cadre du projet Textométrie ANR-06-CORP-029.*

## Sommaire

1. Présentation.....	1
1. Présupposés.....	1
2. Etat de la description et évolutions.....	2
3. Interprétation des grilles remplies et usages.....	2
4. Rappel de la typologie (description brève).....	3
2. Hyperbase.....	5
3. Lexico 3.....	7
4. Grille vierge pour la description d'un autre logiciel.....	9

## 1. Présentation

### 1. *Présupposés*

Ce document s'appuie sur la typologie présentée dans la communication suivante :

Pincemin, B., Heiden, S., Lay, M.-H., Leblanc, J.-M., Viprey, J.-M.

« Fonctionnalités textométriques : Proposition de typologie selon un point de vue utilisateurs »

soumise aux JADT 2010 (<http://jadt2010.uniroma1.it/>, <http://www.jadt.org/>).

La lecture de ce document est donc ici supposée acquise, pour la compréhension des métafonctionnalités recensées.

Dans ce qui suit, nous voulons apporter une illustration concrète de la grille de métafonctionnalités proposée, en mettant chaque métafonctionnalité en lien avec les réalisations correspondantes dans des logiciels de textométrie et d'analyse textuelle.

## **2. *Etat de la description et évolutions***

La liste des logiciels ainsi décrits n'est pas limitée *a priori*. Nous avons commencé par quelques logiciels très connus et typiques de la textométrie telle que décrite dans l'ouvrage de référence de Ludovic Lebart et André Salem (*Statistique textuelle*, Dunod, 1994). Mais bien évidemment l'illustration n'en sera que plus riche, si la typologie est appliquée à d'autres logiciels : nous avons le projet de compléter ce document en introduisant d'autres descriptions, faites par nous-mêmes, ou proposées par des collègues. Egalement, les logiciels évoluent, et leur description demande à être mise à jour régulièrement. Pour ces deux raisons (ajout de nouveaux logiciels, évolution des logiciels existants), ce document de travail est par nature dynamique :

- vous pouvez y contribuer en contactant l'éditeur (Bénédicte Pincemin) ;
- en l'utilisant il faut bien noter la date de la version consultée (en pied de page), et les indications des sources d'information mobilisées pour chaque description (version du logiciel, édition du manuel,...).

## **3. *Interprétation des grilles remplies et usages***

S'agissant de métafonctionnalités, les fonctionnalités nommées sont généralement une manière parmi d'autres de réaliser la métafonctionnalité, autrement dit ici de répondre à un certain questionnement de l'utilisateur (par exemple, concernant les positions d'unités au fil du corpus, on peut apporter une réponse qualitative visuelle, ou/et quantitative basée sur un score statistique). La réalisation peut être partielle, et plus ou moins poussée : par exemple, le calcul d'une concordance avec peu ou pas de possibilités de tri des contextes. Autrement dit, la lecture d'une grille bien remplie ne remplace pas la lecture de la présentation théorique de la typologie. Il faudrait aussi être très prudent si l'on voulait utiliser cette grille pour comparer la couverture fonctionnelle de plusieurs logiciels, puisqu'elle ne permet pas de distinguer entre des réalisations fonctionnelles minimales et d'autres plus complètes. Attention donc, le nombre de cases remplies ne préjuge pas vraiment de la « puissance » du logiciel : on peut avoir un logiciel qui implémente peu de fonctionnalités mais de façon beaucoup plus complète qu'un logiciel plus généraliste.

L'objectif initial qui a motivé la rédaction du présent document est de fournir des exemples illustratifs d'implémentation des métafonctionnalités de la nouvelle typologie proposée, aidant à mieux comprendre, concrètement, à partir de l'expérience d'un logiciel, en quoi peut consister telle ou telle métafonctionnalité. Mais en pratique, c'est aussi une table de correspondance faisant le lien entre les fonctionnalités nommées dans le logiciel et les métafonctionnalités identifiées dans notre typologie, permettant ainsi d'accéder à divers logiciels avec une vue unifiée et orientée usages. Par exemple, à partir d'un besoin donné, comme celui d'étudier l'irrégularité de répartition d'unités entre différentes parties du corpus, on obtient la ou les différentes fonctionnalités pouvant être mises en œuvre pour y répondre, dans les termes du logiciel considéré. Là encore, il faut être prudent, car on peut aussi avoir une fonctionnalité d'un logiciel qui déborde une métafonctionnalité de la typologie, qui trouve mal sa place dans cette typologie et est « à cheval » sur plusieurs métafonctionnalités : la grille donne donc des pistes pour l'utilisation d'un logiciel, mais elles sont toujours à examiner et à comprendre dans la logique du logiciel lui-même avant de les mettre en œuvre.

#### 4. Rappel de la typologie (description brève)

Grille extraite de :

Pincemin, B., Heiden, S., Lay, M.-H., Leblanc, J.-M., Viprey, J.-M. (2009). Fonctionnalités textométriques : Proposition de typologie selon un point de vue utilisateurs. *Soumission aux 10<sup>es</sup> Journées internationales d'Analyse statistique des Données Textuelles (JADT 2010)*, Rome, 9-11 juin 2010.

<i>Structuration</i>	<i>Méta-fonctionnalités</i>	<i>Description brève</i>
DONNEES		Constitution de l'environnement à partir duquel lancer les calculs
Initialisation		
	<b>Profil</b>	Paramétrage par défaut, identification des traces...
	<b>Interfaçage</b>	Importation : création et chargement d'une base textuelle ; exportation de la base pour utilisation dans d'autres logiciels
	<b>Session</b>	Ouverture et fermeture d'une base textuelle, dans un état mémorisé et avec une archive des traitements précédents
Constructions		Définition de nouveaux objets, consultation et gestion d'objets réutilisables
	<b>Propriétés</b>	A partir des propriétés élémentaires disponibles, construction de nouvelles propriétés descriptives, par fusion de valeurs, croisement, etc.
	<b>Sélections</b>	Désignation d'un ensemble d'occurrences, en compréhension (moteur de recherche) ou en extension (par sélection sur une visualisation du corpus)
	<b>Scénarios</b>	Macro-traitement intégrant un enchaînement de traitements, pour l'automatiser
Vue courante		Paramètres généraux, exprimant un état de représentation des données
	<b>Fond</b>	Délimitation du sous-corpus étudié et détermination des unités prises en compte dans les calculs (et servant de référence aux calculs statistiques)
	<b>Structure</b>	Partition et regroupements actifs, le cas échéant ; parallélisme éventuel
	<b>Localisations</b>	Constitution et forme de l'indication précisant le positionnement d'une occurrence
	<b>Plan d'analyse</b>	Système d'unités considéré : segmentation, et propriété définissant les types
	<b>Propriétés d'affichage</b>	Propriété(s) utilisée(s) pour représenter les unités au niveau de l'affichage du texte ou d'un résultat de calcul.
	<b>Focus</b>	Unité ou motif sur lequel se centre le traitement, le cas échéant
LECTURE		Accès au texte, traitements textométriques affichant des occurrences
	<b>Texte</b>	Affichage du texte, en rendant compte visuellement de sa structure logique, avec possibilités de navigation
	<b>Vue interne</b>	Affichage de la représentation interne (structures, propriétés) correspondant au texte ou à un extrait
	<b>Extraits</b>	Liste triable d'extraits définis par un critère et localisés
	<b>Concordance</b>	Liste triable des occurrences d'un focus, alignées verticalement, entourées de leur contexte sur une seule ligne, et avec indication de leur localisation
SYNTHESES		Traitements textométriques affichant des types
Relevés		
	<b>Vocabulaire</b>	Liste triable des unités avec leur fréquence (et leurs sous-fréquences si partition active), pour tout le sous-corpus ou correspondant à un focus
	<b>Mesures</b>	Caractérisations quantitatives du corpus (tailles, gamme des fréquences, indices évaluatifs synthétiques)
Positions		Caractérisation des localisations d'unités dans le corpus
	<b>Déroulement</b>	Disposition et régularité des positions d'unités au fil du texte.
	<b>Distribution</b>	Répartition contrastée d'unités dans les parties d'une partition
	<b>Evolution</b>	Répartition d'unités dans un corpus à partition chronologique ou sérielle

Associations	Liens (syntagmatiques ou non) entre éléments de même nature : plusieurs unités, ou plusieurs parties, ou plusieurs propriétés.	
	<b>Séquences</b>	Repérage d'enchaînements d'unités récurrents
	<b>Cooccurrences</b>	Repérage de la coprésence d'unités au sein d'un même contexte
	<b>Analogies</b>	Repérage de similarités entre unités ou parties, configuration du corpus (par analyse factorielle par exemple), classification
	<b>Corrélations</b>	Force et orientation du lien entre deux propriétés, caractérisées par leurs valeurs au fil des occurrences
ANALYSE	Conduite de l'analyse : qualification des résultats, progression des traitements	
Gestion de la charge	Nécessaire car les calculs textométriques peuvent être complexes et les données et résultats volumineux	
	<b>Progression du calcul</b>	Indicateur de progression du calcul (jauge), possibilité d'interruption (annulation)
	<b>Volume des résultats</b>	Information et possibilités de réduction (échantillonnages,...)
Visualisations	Choix d'une sémiotique de présentation et outils d'exploitation adaptés	
	<b>Tableaux</b>	Transformations sur les lignes et les colonnes (déplacement, suppression, fusion), sur les valeurs (recodage, normalisation...)
	<b>Graphiques</b>	Courbe, diagramme en bâtons, histogramme, camembert... Aide à la lecture, styles d'édition,...
	<b>Diagrammes</b>	Graphes, dendogrammes ; aide à la lecture, styles d'édition...
	<b>Cartes</b>	Aide à la lecture, styles d'édition, projections...
Organisation	Heuristique de mise en forme des résultats accompagnant leur interprétation	
	<b>Filtrage, regroupements, tris</b>	Possibilité de traduire visuellement une organisation des résultats au fur et à mesure de leur analyse
	<b>Synopse et comparaison</b>	Mise en regard de plusieurs résultats et outils d'aide à leur comparaison (projection, différence, tracé de relations...)
Annotation		
	<b>Commentaire</b>	Commentaire libre, signet ; non destiné à l'analyse par des calculs (non indexé pour cela)
	<b>Edition du corpus</b>	Modification du corpus (valeur d'une propriété, segmentation...)
Archive		
	<b>Journal</b>	Systématique – note les fonctions appelées et les paramètres – pour lecture et pour élaboration de scénarios
	<b>Rapport</b>	Edition sélective et documentée de résultats et export (dont impression)

## 2. Hyperbase

<i>Date</i>	<i>Contributeur</i>	<i>Base de la description. Notes éventuelles.</i>
2009.11	B. Pincemin	Logiciel version 8.0 windows. Manuel (Brunet 2006).

Concepteur ou autre contact pour le logiciel : Etienne Brunet

Sites (présentation, téléchargement, développement...) :

<http://www.unice.fr/bcl/spip.php?rubrique38>

### Références bibliographiques

Brunet E. (2006) – *Hyperbase, Logiciel documentaire et statistique pour la création et l'exploitation de bases hypertextuelles, Manuel de référence*, Institut de linguistique française, “Bases, Corpus et Langage”, Université de Nice, mai 2006, 151 pages.

<i>Structuration</i>	<i>Méta-fonctionnalités</i>	<i>Fonctionnalités correspondantes dans le logiciel</i>
DONNEES		
Initialisation		
	<b>Profil</b>	(Installer)
	<b>Interfaçage</b>	Créer une base, Exporter, Anar, Alceste [et certains fichiers aux noms prédéfinis édités par les traitements ; histogr1.xl, histogr2.xl, texte.xl, table.xl, concord.txt, contexte.txt, etc.]]
	<b>Session</b>	(Choix des bases) (Vider)
Constructions		
	<b>Propriétés</b>	
	<b>Sélections</b>	Forme, Expression, Cooccurrents, Initiale, Finale, Chaîne, Lemme, Codes = Codes grammaticaux, Bicode, Tricode, Syntaxe=Structure syntaxique, Liste, Fichier, Segments répétés [+ clic sur les unités dans certains affichages comme Lecture ou Index]
	<b>Scénarios</b>	
Vue courante		
	<b>Fond</b>	Choix du corpus
	<b>Structure</b>	Réinitialiser, Liste/Genre
	<b>Localisations</b>	Réinitialiser
	<b>Plan d'analyse</b>	[propriété d'analyse multiple (en parallèle sur toutes les propriétés), ou déterminé par le mode de construction du focus, et par détail/cumul]
	<b>Propriétés d'affichage</b>	[identique à la propriété d'analyse]
	<b>Focus</b>	Liste [peut jouer le rôle de focus persistant] [+ cf. Sélections]
LECTURE		
	<b>Texte</b>	Lecture, (Mot)
	<b>Vue interne</b>	(Lecture)
	<b>Extraits</b>	Contexte, Phrases-clés = Passages spécifiques
	<b>Concordance</b>	Concordance
SYNTHESES		
Relevés		
	<b>Vocabulaire</b>	Index, Mot, Liste, Index hiérarchique = Hautes fréquences, Edition/Créer/ =Concordance/Tout/ : Index, Dictionnaire
	<b>Mesures</b>	Distribution/ : Etendue et probabilités, Richesse et hapax, Distribution des fréquences, Pareto ; Liste/ : Catégorie, Fréquence, Groupe, Longueur, Lettres, Indices(=Champs ?)

Positions	
<b>Déroulement</b>	Evolution ; Distribution/ : Evolution alphabétique, Evolution hiérarchique, Tranches ; Topologie
<b>Distribution</b>	Graphique, Spécificités, (Lecture/Ecart, Phrases-clés = Passages spécifiques)
<b>Evolution</b>	Distribution/ : Croissement chronologique, Croissement inverse
Associations	
<b>Séquences</b>	Segments répétés
<b>Cooccurrences</b>	Contexte/Thème ; Associations/, Thèmes/, Corrélats/ : Graphe, Histogramme
<b>Analogies</b>	Factorielle, Arbre=Analyse arborée, Distribution/Distance, Liste/Connexion, Alceste, Graphique/Double
<b>Corrélations</b>	
ANALYSE	
Gestion de la charge	
<b>Progression du calcul</b>	[bas de la fenêtre]
<b>Volume des résultats</b>	
Visualisations	
<b>Tableaux</b>	Liste
<b>Graphiques</b>	Graphique, Distribution/Histogramme
<b>Diagrammes</b>	Arborée/Cacher les noms, Cacher les chiffres, Graphe (Radiale/Rectangulaire), Taille, Echelle [déplacer un point ou un nœud]
<b>Cartes</b>	Factorielle/Voir les chiffres, Affiche, Taille, Zoom, Changer la couleur d'un point, [déplacer un point]
Organisation	
<b>Filtrage, regroupements, tris</b>	Légende ; Arborée/ : Figure, Mettre noms en couleurs, Chronologie ; Factorielle/ : Ovale, Ligne, Colorier, Changer la couleur d'un point
<b>Synopse et comparaison</b>	(Lecture/Ecart, Factorielle/Affiche, Arborée/ : Mettre noms en couleurs, Chronologie)
Annotation	
<b>Commentaire</b>	(Biblio)
<b>Edition du corpus</b>	
Archive	
<b>Journal</b>	
<b>Rapport</b>	Notes, Edition, Imprimer

### 3. Lexico 3

<i>Date</i>	<i>Contributeur</i>	<i>Base de la description (sources). Notes éventuelles.</i>
2009.11	B. Pincemin	Logiciel version 3.45.0.1. Manuel (Fracchiolla <i>et al.</i> , 2003).

Concepteur ou autre contact pour le logiciel : André Salem

Sites (présentation, téléchargement, développement...) :

<http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/index.htm>

#### Références bibliographiques

Fracchiolla B., Kuncova A., Maisondieu A. (2003) – *Lexico 3, outils de statistique textuelle, Manuel d'utilisation*, Version 3.41, SYLED-CLA2T, Université Paris 3, février 2003, 50 pages.

<i>Structuration</i>	<i>Méta-fonctionnalités</i>	<i>Fonctionnalités correspondantes dans le logiciel</i>
DONNEES		
Initialisation		
	<b>Profil</b>	
	<b>Interfaçage</b>	Nouveau
	<b>Session</b>	Ouvrir
Constructions		
	<b>Propriétés</b>	
	<b>Sélections</b>	[sélection par désignation (clic souris)] Groupe de formes=TGen/: exactement, début, fin, contient, expression rationnelle, Ajouter, Supprimer ; [Enregistrement :] Groupe de formes, Garde-mots
	<b>Scénarios</b>	
Vue courante		
	<b>Fond</b>	PCLC principales caractéristiques lexicométriques
	<b>Structure</b>	Statistiques par partie, PCLC Principales caractéristiques lexicométriques
	<b>Localisations</b>	
	<b>Plan d'analyse</b>	
	<b>Propriétés d'affichage</b>	
	<b>Focus</b>	[saisie dans champ spécifique à la fonctionnalité, ex.:] Forme [drag and drop depuis Dictionnaire (et autres listes onglets du cadre gauche) ou Groupe de formes ou Garde-mots ; sélection de parties dans tableau PCLC ou dans carte AFC]
LECTURE		
	<b>Texte</b>	TextPloreur
	<b>Vue interne</b>	
	<b>Extraits</b>	(Carte des sections)
	<b>Concordance</b>	Concordance
SYNTHESES		
Relevés		
	<b>Vocabulaire</b>	Dictionnaire, Groupe de formes
	<b>Mesures</b>	PCLC Principales caractéristiques lexicométriques
Positions		
	<b>Déroulement</b>	Carte des sections, PCLC/Accroissement du vocabulaire
	<b>Distribution</b>	Statistiques par partie/Graphique de ventilation, PCLC/: Spécificités, Spécificités totales ; (Carte des sections/Spécificités)
	<b>Evolution</b>	Spécificités diachroniques/: Spécificités évolutives, Spécificités

		chronologiques
Associations	<b>Séquences</b>	Segments répétés, Concordance/Inventaire distributionnel
	<b>Cooccurrences</b>	Carte des sections/Cooccurrences
	<b>Analogies</b>	AFC
	<b>Corrélations</b>	
ANALYSE		
Gestion de la charge	<b>Progression du calcul</b>	
	<b>Volume des résultats</b>	
Visualisations	<b>Tableaux</b>	
	<b>Graphiques</b>	
	<b>Diagrammes</b>	Carte des sections/ : [échelle], Seuillage
	<b>Cartes</b>	AFC/ : Valeurs propres, [coordonnées]
Organisation	<b>Filtrage, regroupements, tris</b>	AFC/[possibilité de colorier les points]
	<b>Synopse et comparaison</b>	[multifenêtrage et « feuilles de travail »] Créer une nouvelle feuille, Déplacer vers une autre feuille, Mosaïque ; [usage des couleurs pour cumuler plusieurs vues (Ventilation, Carte des sections)]
Annotation	<b>Commentaire</b>	
	<b>Edition du corpus</b>	
Archive	<b>Journal</b>	Navigation, Liste des fenêtres
	<b>Rapport</b>	Ajouter au rapport



#### 4. Grille vierge pour la description d'un autre logiciel

<i>Date</i>	<i>Contributeur</i>	<i>Base de la description (sources). Notes éventuelles.</i>
aaaa.mm		Logiciel version ..., références bibliographiques. <i>Commentaires éventuels.</i>

Concepteur ou autre contact pour le logiciel :

Sites (présentation, téléchargement, développement...) :

http://

#### *Références bibliographiques*

<i>Structuration</i>	<i>Méta-fonctionnalités</i>	<i>Fonctionnalités correspondantes dans le logiciel</i>
<b>DONNEES</b>		
Initialisation		
	<b>Profil</b>	
	<b>Interfaçage</b>	
	<b>Session</b>	
Constructions		
	<b>Propriétés</b>	
	<b>Sélections</b>	
	<b>Scénarios</b>	
Vue courante		
	<b>Fond</b>	
	<b>Structure</b>	
	<b>Localisations</b>	
	<b>Plan d'analyse</b>	
	<b>Propriétés d'affichage</b>	
	<b>Focus</b>	
<b>LECTURE</b>		
	<b>Texte</b>	
	<b>Vue interne</b>	
	<b>Extraits</b>	
	<b>Concordance</b>	
<b>SYNTHESES</b>		
Relevés		
	<b>Vocabulaire</b>	
	<b>Mesures</b>	
Positions		
	<b>Déroulement</b>	
	<b>Distribution</b>	
	<b>Evolution</b>	
Associations		
	<b>Séquences</b>	
	<b>Cooccurrences</b>	
	<b>Analogies</b>	
	<b>Corrélations</b>	

<b>ANALYSE</b>		
Gestion de la charge	<b>Progression du calcul</b>	
	<b>Volume des résultats</b>	
Visualisations	<b>Tableaux</b>	
	<b>Graphiques</b>	
	<b>Diagrammes</b>	
	<b>Cartes</b>	
Organisation	<b>Filtrage, regroupements, tris</b>	
	<b>Synopse et comparaison</b>	
Annotation	<b>Commentaire</b>	
	<b>Edition du corpus</b>	
Archive	<b>Journal</b>	
	<b>Rapport</b>	