

Mode d'emploi de l'interface guidée Scientext

Achille Falaise, Agnès Tutin

04/05/2010

scientext@u-grenoble3.fr

L'utilisation de l'interface simple et guidée se fait en trois temps :

1. Sélection du corpus
2. Recherche des expressions sélectionnées
3. Affichage et statistiques

Ces étapes sont détaillées dans les sections suivantes.

1. Etape 1 : Sélection du corpus

Une fois sur la page des textes, Sélectionner ensuite les critères pertinents : discipline (ou famille de disciplines), genre textuel, partie textuelle.

Disciplines	Genres	Parties
<input checked="" type="checkbox"/> Sciences humaines	<input checked="" type="checkbox"/> Article	<input checked="" type="checkbox"/> Parties principales
<input checked="" type="checkbox"/> Linguistique	<input checked="" type="checkbox"/> Communication	<input checked="" type="checkbox"/> Développement
<input checked="" type="checkbox"/> Psychologie	<input checked="" type="checkbox"/> Thèse	<input checked="" type="checkbox"/> Introduction
<input checked="" type="checkbox"/> Sciences de l'éducation	<input checked="" type="checkbox"/> HDR	<input checked="" type="checkbox"/> Conclusion
<input checked="" type="checkbox"/> Traitement Automatique des Langues		<input checked="" type="checkbox"/> Autres parties
<input checked="" type="checkbox"/> Sciences expérimentales		<input checked="" type="checkbox"/> Résumé
<input checked="" type="checkbox"/> Biologie		<input checked="" type="checkbox"/> Notes
<input checked="" type="checkbox"/> Médecine		<input checked="" type="checkbox"/> Remerciements
<input checked="" type="checkbox"/> Sciences appliquées		<input checked="" type="checkbox"/> Annexe
<input checked="" type="checkbox"/> Électronique		<input checked="" type="checkbox"/> Titres
<input checked="" type="checkbox"/> Mécanique		
<input type="button" value="Tout"/> <input type="button" value="Rien"/>	<input type="button" value="Tout"/> <input type="button" value="Rien"/>	<input type="button" value="Tout"/> <input type="button" value="Rien"/>

Aller ensuite sur Suite. La liste des textes sélectionnés s'affiche. On peut effectuer le tri par discipline, type de texte, titre, recueil et auteur. On pourra cocher les textes que l'on souhaite interroger.

Pour passer à la recherche dans les textes, on cliquera sur Suite.

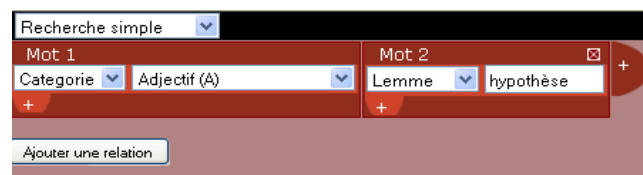
2. Recherche guidée dans les textes

Trois modes de recherche sont prévus pour explorer les textes :

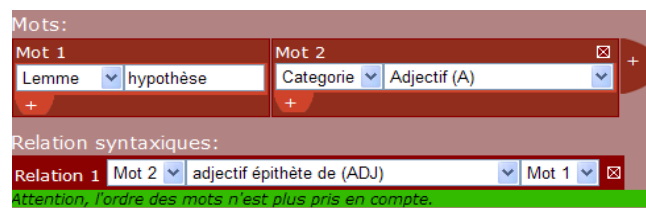
- un mode de recherche sémantique, avec des grammaires prédéfinies (pas détaillé ici).
- un mode simple par recherche guidée.
- un mode de recherche avancé (qui ne sera pas détaillé ici) et qui utilise des expressions régulières.

C'est ici le mode de recherche guidée qui est décrit.

L'utilisateur utilise ici des menus, qui indiquent pour les mots : **la forme, la catégorie syntaxique** ou **le lemme**. Il est aussi possible d'utiliser des relations syntaxiques, grâce à une analyse syntaxique réalisée automatiquement à l'aide du logiciel Syntex (de Didier Bourigault). Ces possibilités apparaissent dans des ascenseurs. La recherche suivante permet de repérer les suites de mots où un adjectif apparaît avant le lemme *hypothèse*. En revanche, la recherche ne prendra pas en compte les cas où l'adjectif est postposé au nom ou les cas où l'adjectif n'est pas immédiatement antérieur au nom.



Pour proposer une requête plus adaptée, il faudra utiliser les relations syntaxiques (pour cela, cliquer sur « Ajouter une relation »). On indique alors la relation syntaxique (de dépendance) entre les mots. **Attention, dans ce cas, il n'y a plus d'ordre linéaire !** La recherche suivante extrait les adjectifs épithètes qui accompagnent *hypothèse* indépendamment de leur position.



Les recherches qui utilisent les relations syntaxiques sont généralement plus intéressantes que celles qui n'utilisent que l'ordre linéaire, mais elles sont un peu plus complexes à maîtriser.

Pour les recherches, il est possible de paramétrer le nombre de résultats voulus. Par défaut, la taille est limitée à 20 réponses. Il est conseillé d'indiquer une limite, car le corpus est assez volumineux.

Pour obtenir les résultats sous forme de concordance, il suffit de cliquer sur Suite.

3. Affichage des résultats et résultats

3.1 Affichage

Par défaut, l'affichage des résultats se fait dans une concordance KWIC dont la taille est paramétrable (par défaut, 10 mots avant, 10 mots après dans la même phrase), comme ci-dessous. Un code indique le genre du texte et la discipline.

20 occurrences.		Page: 1		
N°	Contexte gauche: 10 mots	Séquence recherchée:	Contexte droit: 10 mots	Réf. texte
1	complexes métalliques , cette théorie se base sur les	hypothèses suivantes :		[ele-the-151-body]
2	ce jour , pas de confirmations expérimentales à ces	deux hypothèses	[35] .	[ele-the-151-body]
3	de ces souches a été nécessaire pour trancher entre les	deux hypothèses	.	[bio-the-34-body]
4	colonisés et traités par ITZ a déjà suggéré une	telle hypothèse	.	[bio-the-34-body]
5		De telles hypothèses	impliquent que l' acte de langage expressif est un acte	[lin-com-120-body]
6	Au final [25] , l'	hypothèse générale	d' une influence plus forte de la discipline que de la	[lin-the-27-body]
7	selon qu' elles entendent multiplier les dimensions prises en compte sans	hypothèse préalable	sur les caractéristiques du genre [27] , ou	[lin-the-27-body]
8		Une seconde hypothèse	intervient , pour étudier les pratiques d' écriture de l'	[lin-the-27-body]
9	article vise à asseoir dans la matérialité du texte l'	hypothèse rhétorique	que le genre implique un rôle relativement stable endossé par	[lin-the-27-body]
10	Une des	principales hypothèses	de notre recherche est que le genre , parce qu' il	[lin-the-27-body]
11	nous avons dit que nous admettions la validation de l'	hypothèse non-nulle	(autrement dit l' hypothèse d' un effet significatif)	[lin-the-27-body]
12	le logiciel , nous avons retenu 97 variables par	hypothèses pertinentes	pour une analyse du genre basée sur l' unité phrase	[lin-the-27-body]

L'utilisateur peut avoir un contexte plus large en cliquant sur la concordance, comme ci-dessous.

1	complexes métalliques , cette théorie se base sur les	hypothèses suivantes :	[ele-the-151]
[the-151] Maylis ORIO - <i>Etude par la Théorie de la Fonctionnelle de la Densité des propriétés électroniques et magnétiques complexes de fer. Application aux systèmes de types Catalase et Fer-Soufre</i>			
<p>en compte un terme supplémentaire d'énergie potentielle ce qui nous amène à considérer l'hamiltonien de travail suivant :</p> <p>(2.18)</p> <p>Dans cette équation, le premier terme contient les énergie cinétique et potentielle du système d'électrons ainsi considéré, le second terme fait référence à la répulsion électronique et le troisième terme correspond à l'interaction de spin-orbite. Le dernier terme va, quant à lui, traduire l'interaction entre un ion libre de charge e et son environnement représenté par N charges négatives Z_i, r_{ij} étant à la distance entre l'ion considéré et chacune des N charges.</p> <p>Cette approche est nommée "théorie du champ cristallin" et s'est avérée très satisfaisante pour évaluer les effets d'environnement sur les niveaux d'énergie des ions. Cette théorie a été formulée par Bethe en 1929 [8] et correspond à un modèle électrostatique décrivant des sous-couches d ou f de l'atome dans des entourages de diverses symétries.</p> <p>Dans le cas des complexes métalliques, cette théorie se base sur les hypothèses suivantes :</p>			

Il est également possible d'afficher l'analyse syntaxique de la phrase correspondant à la requête, en cochant Voir l'analyse syntaxique. Par exemple, la figure ci-dessous indique l'analyse syntaxique pour la phrase *Dans le cas des complexes métalliques, cette théorie se base sur les hypothèses suivantes* :

☒ Voir l'analyse syntaxique

Dépendance ☒

Lemme ☐

Forme ☒

Catégorie ☐

Dans le cas des

NOMPREP ☐ complexes

ADJ ☐ métalliques

,

base

SUJ ☐ théorie

DET ☐ cette

REF ☐ se

PREP ☐ sur

NOMPREP ☐ hypothèses

DET ☐ les

ADJ ☐ suivantes

:

Il est possible d'extraire le résultat des concordances en format html ou CSV (voir au bas de la fenêtre).

3.2 Statistiques

On peut avoir l'affichage de quelques statistiques simples. Pour cela, il faut cliquer sur statistiques ou sur Suite à partir des concordances.

On peut ainsi obtenir :

- la liste des lemmes correspondant à la requête, comme ci-dessous.

Quelles statistiques souhaitez-vous consulter ?

Liste des lemmes

Lemme	occurrences
/hypothèse/ /général/	34
/hypothèse/ /suivant/	23
/hypothèse/ /alternatif/	23
/hypothèse/ /théorique/	18
/hypothèse/ /explicatif/	15
/hypothèse/ /principal/	14
/hypothèse/ /nul/	12
/hypothèse/ /interprétatif/	12
/hypothèse/ /émotionnel/	10
/hypothèse/ /local/	9
/hypothèse/ /issu/	9
/hypothèse/ /central/	6
/hypothèse/ /neuropsychologique/	6
/hypothèse/ /linguistique/	6
/hypothèse/ /initial/	5
/hypothèse/ /préalable/	4
/hypothèse/ /précédent/	4

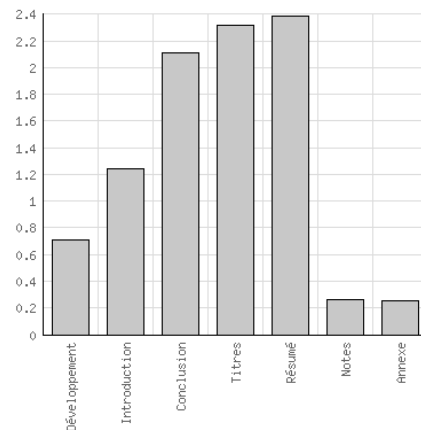
- la répartition des lemmes (fréquences absolues, relatives et schémas) selon :
 - o la partie textuelle (Cf Schéma ci-dessous).
 - o la discipline
 - o le type de texte

- le texte

Quelles statistiques souhaitez-vous consulter ?

Répartition des lemmes

Partie textuelle	Nombre absolu d'occurrences	Nombre de mots total	Nombre relatif d'occurrences
Développement	258 /	3645846 =	0.71 ‰
Introduction	26 /	209148 =	1.24 ‰
Conclusion	18 /	85205 =	2.11 ‰
Titres	13 /	56121 =	2.32 ‰
Résumé	6 /	25061 =	2.39 ‰
Notes	4 /	153386 =	0.26 ‰
Annexe	3 /	118200 =	0.25 ‰



Il est également possible de rapatrier les résultats au format CSV pour une utilisation avec un tableur.