

Concordances et concordanciers

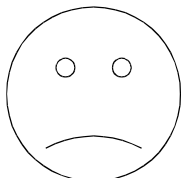
De l'art du bon KWAC

Bénédicte PINCEMIN
CNRS & Université de Paris 13, LLI

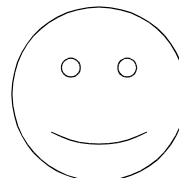
*« Corpus en Lettres et Sciences sociales :
des documents numériques à l'interprétation »
Colloque international et école d'été, Albi, 10-14 juillet 2006*

Du nouveau ??

- Une technique bien connue
- Un traitement (automatique) peu élaboré

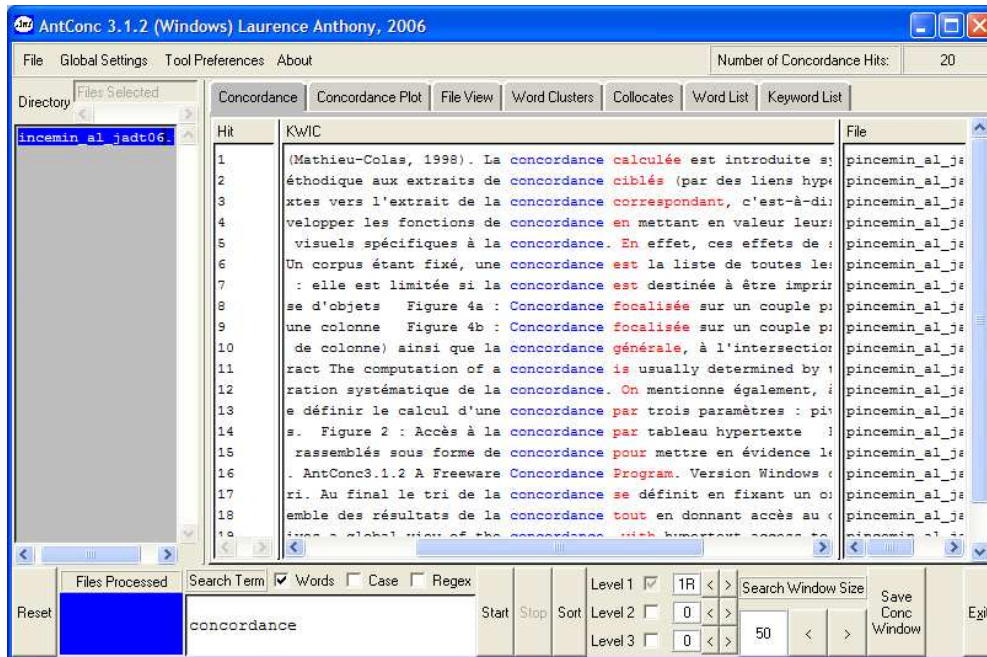


Les gens sont
simplistes...



Pourquoi ça marche ?
Comment ça marche ?

Exemple de concordancier : AntConc



Les trois paramètres des concordanciers

- 1) Pivot : mot ou autre item linguistique (lemme, expression,...)
- 2) Taille de contexte
- 3) Tri :
 - corpus,
 - pivot,
 - 1G (1er mot qui précède – « à gauche »),
 - 1D (1er mot qui suit – « à droite »)

L'atout des concordances

Tri + alignement vertical en colonne

= superposition et répétitions

= effets visuels, heuristique de lecture et d'analyse

Exemples de KWAC

⇒ Le mot (centré) puis le contexte = KWAC

vs le contexte puis le mot = KWIC

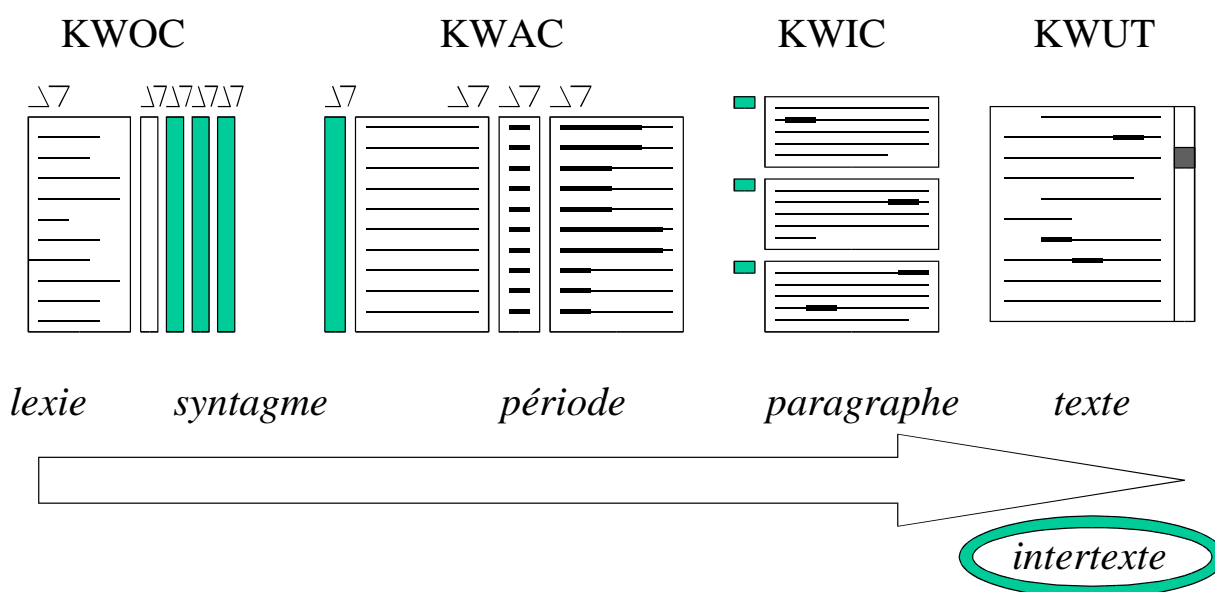
Définition du (bon) KWAC

- Un **corpus** étant fixé,
- Liste de toutes les occurrences d'un **pivot**
- Alignées verticalement en **colonne**,
- Entourées de part et d'autre par leur **contexte**
- sur une **ligne**,
- Accompagnées d'une **référence** indiquant de façon pertinente leur positionnement dans le corpus,
- Et **triées** selon un critère pertinent pour l'analyse.

Typologie des relevés d'occurrence : formulation mnémotechnique

- KWOC = KeyWord **O**ut of Context
- KWAC = KeyWord **A**nd Context = concordance (heuristique visuelle = contexte sur une ligne + colonne + tri)
- KWIC = KeyWord **I**n Context
- KWUT = KeyWord **U**p to Text

Allure des KWOC-KWAC-KWIC-KWUT



Logiciels de Textométrie

	KWOC	KWAC	KWIC	KWUT
Hyperbase	<i>(Liste, Mot)</i>	<i>Concordance</i>	<i>Contexte</i>	Lien hypertexte depuis <i>Mot</i>
Lexico 3	<i>Groupes de for</i>	<i>Concordances</i>	(<i>Carte des s</i>)	<i>Textploreur, et</i> <i>Carte des sections</i>
Weblex	<i>Index [html 1]</i> <i>[html 2]</i>	<i>Concordance,</i> option « sur une ligne » <i>[html]</i>	<i>Contextes</i> <i>[html]</i>	(Lien hypertexte depuis référence dans <i>Concordance</i> ou <i>Contextes [html]</i>)

Typologie des relevés d'occurrences

	KWOC	KWAC	KWIC	KWUT
	Out	And	In	Up to
Ex.	<i>is ... -ed</i> <i>protect-</i>	<i>shall (tri droit)</i>	<i>lecture</i>	<i>création+</i>
<i>Corpus</i>	<i>Droits de l'Homme</i>	<i>Droits de l'Homme</i>	<i>Littérature</i>	<i>Bible</i>
<i>Logiciel</i>	Weblex	Weblex	Hyperbase	Lexico 3
Atout	Synthétique Pics et lacunes	Heuristique visuelle par alignement (superposition) et tri (répétition)	Optimisation des contextes	Répartition globale et locale

KWOC dans Hyperbase

The screenshot shows the Hyperbase interface with a list of words on the left and a detailed KWOC analysis for the word 'lecture' on the right. The analysis includes a list of words, their frequencies, and a table of word forms.

liste de mots

1 Lecoureur, 81
28 lecteur, 627382 111124 202211 2213
25 lecteurs, 697481 102131172 201225
3 lectrice, 213
75 lecture, 1121314154688692 103136145174181 19120121182212
32 lectures, 215882

fréquence totale: 75

lecture
CLIQUEZ SUR UN TEXTE (ou sur TOUS) pour y repérer les contextes du mot "lecture"
Cliquez AILLEURS

Liste de mots

Effacer un mot: CLIC + MAJ
GRAPHIQUE: clic sur un mot ou un texte

lecteur	lecteurs	lectrice	lecture	lectures	livres
0	0	0	0	0	0
0	0	0	2	1	13
0	0	0	0	9	4
0	2	0	0	1	0
0	0	0	0	0	5
0	0	0	0	0	25
0	0	0	0	0	0
0	0	0	0	3	0
1	1	1	1	4	8
0	4	1	1	1	18
0	0	1	0	0	8
1	1	0	0	0	1
2	1	2	13	11	9
3	3	0	56	1	26

Typologie

Logiciels

KWOC dans Lexico 3

The screenshot shows the Lexico 3 interface with a list of words on the left and a detailed KWOC analysis for the word 'création' on the right. The analysis includes a list of words, their frequencies, and a table of word forms.

Graphique de ventilation pour la partition : Livre

01_Genèse 08_Ruth 15_Esdras 23_Esaie 29_Joel 36_Sophonie 4

Formes (ordre lexicographique) | Fréquence

Forme	Fréquence
crachements	1
crayon	1
crèbe	11
crèbe	1
crèancier	5
crèanciers	3
crèateur	14
crèation	14
crèature	7
crèatures	4
crèche	7
crèches	3
crèe	37
crèe	4
crèées	4
crèer	6
crèera	1
crèèrent	1
crèés	7
crème	10
crèreaux	2
crèpuscule	5
crèusa	6
crèusait	1
crèuse	8
crèusé	10
crèusée	3
crèusées	3
crèusent	1
crèuser	1
crèuserai	1
crèuserent	4
crèuses	1
crèusés	2
crèuset	10

TextPloreur

<Testament=Ancien>
<Genre=Loi>
<Livre=01_Genèse>
§<Verset=Gn1:1> Au commencement, Dieu créa les cieux et la terre.
§<Verset=Gn1:2> La terre était informe et vide: il y avait des ténèbres, et l'esprit de Dieu se mouvait au-dessus des eaux.
§<Verset=Gn1:3> Dieu dit: Que la lumière soit! Et la lumière fut.
§<Verset=Gn1:4> Dieu vit que la lumière était bonne; et Dieu bénit le jour, et dit: Que le jour s'appelle jour, et la nuit s'appelle nuit.
§<Verset=Gn1:5> Dieu appela la lumière jour, et la nuit s'appela nuit. Et Dieu termina son ouvrage le sixième jour.
§<Verset=Gn1:6> Dieu dit: Qu'il y ait une étendue entre les cieux et la terre.

Groupes de formes

Nom du groupe: création
Le motif: est exactement ce que je recherche
 Ajouter
 Rechercher
 Enregistrer
 Supprimer
 Charger

Forme	Fréquence
créateur	14
création	14
créa	11
créature	7
créé	7
créés	7
créer	6
créatures	4
créée	4
créées	4
créai	1
créata	1
créèrent	1

Typologie

Logiciels

KWOC dans Weblex

Index des occurrences de [word="is"] [0,2] [word=".*ed"] dans le corpus dh-lang-EN

ord	f	événement
1	13	is entitled
2	7	is satisfied
3	4	is deposited
4	4	is provided
5	3	is based
6	3	is deprived
7	3	is designated

Corpus: dh-lang-EN
Source: [word="is"] [0,2] [word=".*ed"]

(CQP : Index Concordances Références Contextes Répartition Spécificités)

Vocabulaire Répartitions Pareto Zipf Longueur des phrases Dimensions
Lexicogramme Lexicogramme récursif Lexicogrammes récursifs Cooccurrences
Séquences rénétiées Termes

Général
Index
Concordances
Vocabulaire

Options générales de l'affichage

Index des occurrences de protect.* dans le corpus dh/lang EN=geo-amerique

ord	f	événement
1	22	protection
2	5	protect
3	2	protected
4	2	protecting
5	1	protective
	32	au Total

Index des occurrences de protect.* dans le corpus dh/lang EN=geo-europe

ord	f	événement
1	82	protection
2	5	protect
3	3	protected
	90	au Total

Index des occurrences de protect.* dans le corpus dh/lang EN=geo-univers

Typologie
Logiciels

KWAC dans Hyperbase

C:\HYPERBAS\EXEMPLE.EXE

Retour Forme Lemme Expr. Initial Final Chain Liste Tout Nb 75 CONCORDANCE

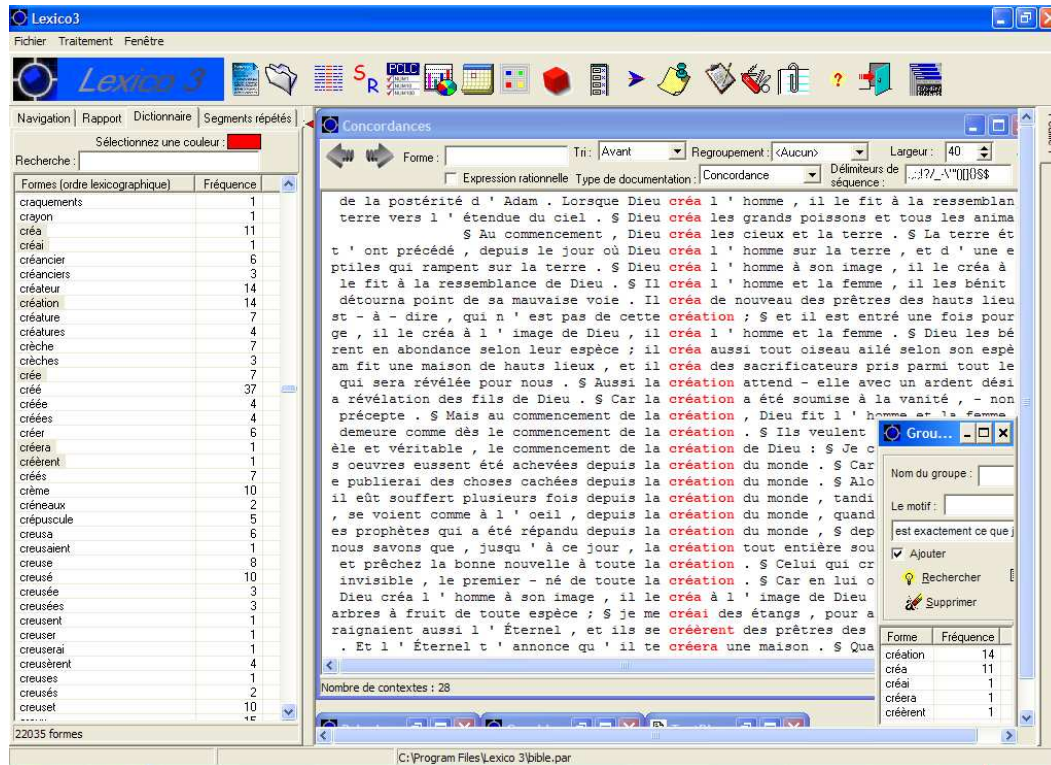
Cliquer sur une ligne pour voir la page

Tm8654b| e une confu sion avec une ancienne lecture , ou un rêve récent ? Le rêv
Po2770b| tentivement ce grimoire . Après lecture il laissa tomber le papier et
Tm8580b| m' aider d' aucune règle , cette lecture consistant en un acte de créat
St7058d| ces ridicules légendes . Cette lecture ne laissa pas de me suggérer q
Ca 453d| ennuyaient autant que moi à cette lecture : tous les gens sincères m' on
Ra1380b| idées , et qu' il y a dans cette lecture un poison subtil , caché sous
Bo3992d| ains , il s' enfonçait dans cette lecture . Tandis qu' il étudiait les
Hé 613b| savez ce que je vous dis de cette lecture et de la conduite du théologie
Hé 567c| que vous . Moins vous aurez de lecture à faire , mieux il faudra la c
Bu4564d| ction et il demanda au cabinet de lecture des romans d' amour .
Bo3814d| toujours abonnée à un cabinet de lecture . - - Si Madame veut me fair
Bu4534d| m , un abonnement à un cabinet de lecture et leur expédia les lettres d'
Bu4707a| ent un abonnement à un cabinet de lecture . Les maîtres anciens étaien
Bo3723a| e poussiè re des vieux cabinets de lecture . Avec Walter Scott , plus t
Ra1522c| u' on ne faisait point à table de lecture pieuse et qu' on jouait aux ca
Em1113a| pensé de l' apprendre . Trop de lecture ne sert qu' à faire de présomp
Tm8342e| ' écrivait pas mieux , en donnait lecture à haute voix .
Ra1684b| abbé Jacques de La Cour leur fit lecture : Dieu , disait - il , connaît
Rq5716a| er bien épais , bien lourd . La lecture lui ouvrit des horizons romane
Sw7354d| ent une inviolable solitude : la lecture , la rêverie , les larmes et l
Sw8086c| e il était convaincu que non ; la lecture de la lettre anonyme en avait
Ma 22b| pprimes l' état du frère . A la lecture de la lettre qui nous en infor
Po2863a| éta pas la moindre attention à la lecture de cette pro curation générale
Bo3979d| . Mais Binet , tout entier à la lecture de l' addition , n' avait rien
Sw7513a| fois par une visite de Swann à la lecture que j' étais en train de faire
Lu6891c| nt , ses yeux se troublèrent à la lecture des vingt mots de ce télégramm
Em1110d| e Spectateur , dont vous aimez la lecture .
Ch1809a| Eh bien , disait Hulot après la lecture publique de ce discours consul
Em1116d| l' art de penser , que , dans la lecture , leur esprit est au moins gui
Sw7503c| ous ne la connaissons que dans la lecture , en imagination : dans la réa
Rq5716b| les yeux . Ce subit amour de la lecture eut une grande influence sur s
Em1169d| ouche . J' attends la fin de la lecture , et puis montrant au gouverne
Sw7508a| lieu pour moi ; l' intérêt de la lecture , magique comme un profond som
Tm8251d| apprend à rabaisser le prix de la lecture , et nous montre que ce que l'

Regroupement des formes d'un même vocable

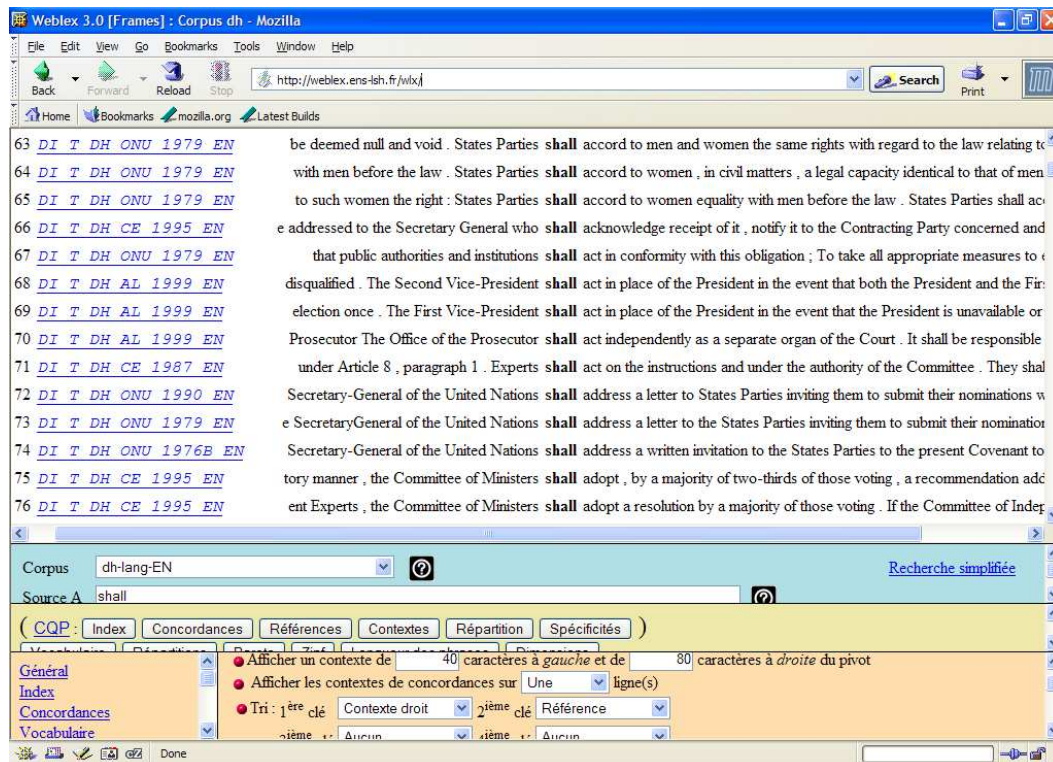
Typologie
Logiciels

KWAC dans Lexico 3



Typologie
Logiciels

KWAC dans Weblex



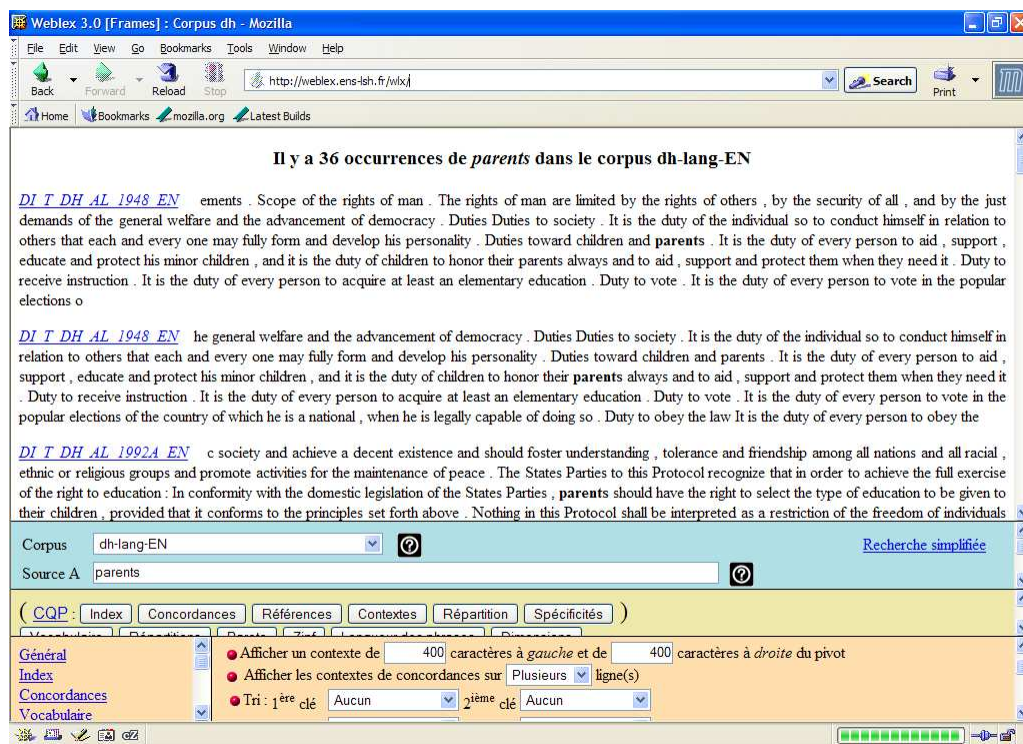
Déf. KWAC
Typologie
Logiciels

KWIC dans Hyperbase



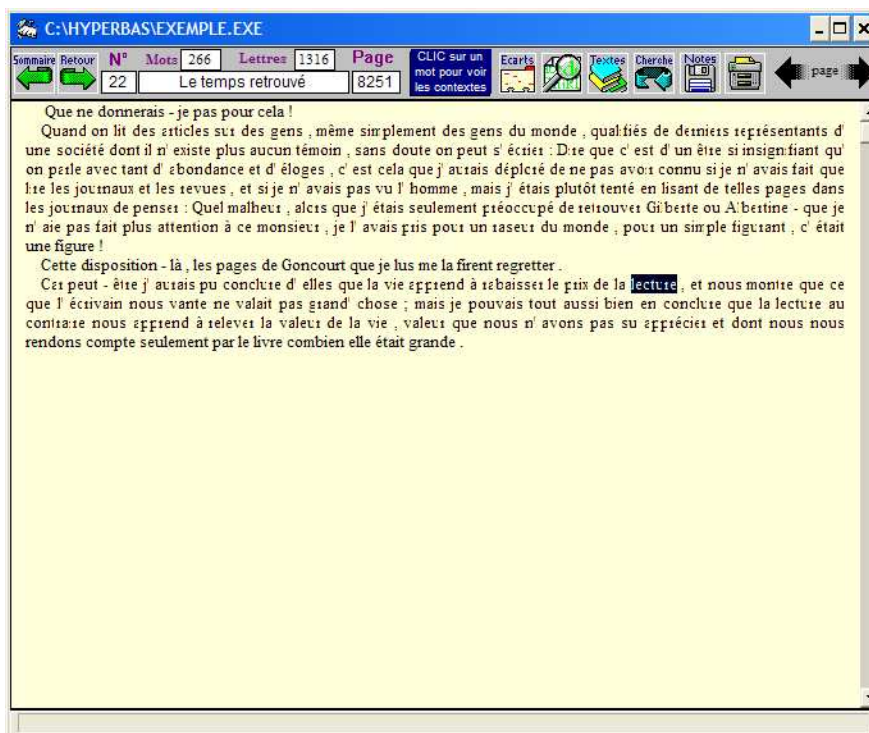
Typologie
Logiciels

KWIC dans Weblex



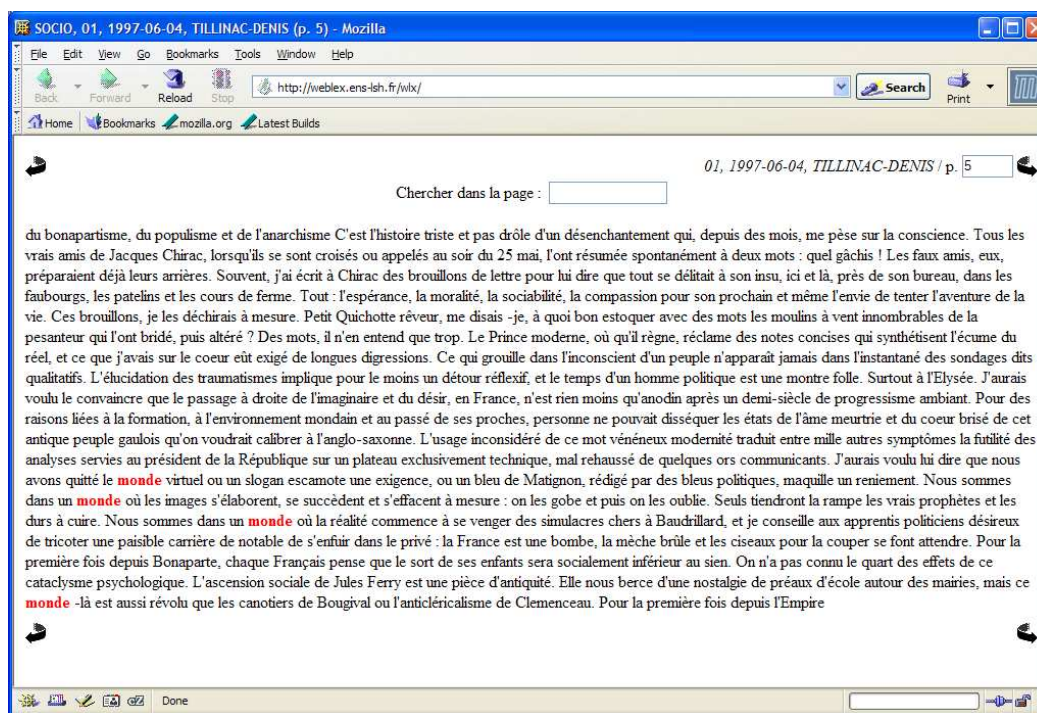
Typologie
Logiciels

KWUT dans Hyperbase



Typologie
Logiciels

KWUT dans Weblex



Typologie
Logiciels

KWUT dans Lexico 3

The screenshot shows the Lexico 3 interface with a search for 'cré'. The 'Groupes de formes' window displays a list of forms and their frequencies:

Forme	Fréquence
créé	37
créateur	14
création	14
créa	11
créature	7
crée	7
créées	7
créés	7
créera	6
créèrent	4
créatures	4
crée	4
créées	4
créai	1
créera	1
créèrent	1

The 'TextPloreur' window shows a KWUT analysis of the word 'création' in a biblical text, with sections highlighted in yellow:

reptiles qui rampent sur la terre .
 §<Verset=Gn1:27> Dieu **créa** l' homme à son image , il le **créa** à l' image de Dieu , il **créa** l' homme et la femme .
 §<Verset=Gn1:28> Dieu les bénit , et Dieu leur dit : Soyez féconds , multipliez , remplissez la terre , et l' assujettissez ; et dominez sur les poissons de la mer , sur les oiseaux du ciel , et sur tout animal qui se meut sur la terre .
 §<Verset=Gn1:29> Et Dieu dit : Voici , je vous donne toute herbe portant de la semence et qui est à la surface de toute la terre , et tout arbre ayant en lui du fruit d' arbre et portant de la semence : ce sera votre nourriture .
 §<Verset=Gn1:30> Et à tout animal de la terre , à tout oiseau du ciel , et à tout ce qui se meut sur la terre , ayant en soi un souffle de vie , je donne toute herbe verte pour nourriture . Et cela fut ainsi .
 §<Verset=Gn1:31> Dieu vit tout ce qu' il avait fait et voici , cela était très bon . Ainsi , il y eut un soir , et il y eut un matin : ce fut le sixième jour .
 §<Verset=Gn2:1> Ainsi furent achevés les **cieux** et la terre , et toute leur armée .
 §<Verset=Gn2:2> Dieu acheva au septième jour son oeuvre , qu' il avait faite : et il se reposa au septième jour de toute son oeuvre , qu' il avait faite .
 §<Verset=Gn2:3> Dieu bénit le septième jour , et il le sanctifia , parce qu' en ce jour il se reposa de toute son oeuvre qu' il avait **créée** en la faisant .
 §<Verset=Gn2:4> Voici les origines des **cieux** et de la terre , quand ils furent **créés** . Lorsque l' Éternel Dieu fit une terre et des **cieux** , §<Verset=Gn2:5> aucun arbuste des champs n' était encore sur la terre . et aucune herbe des champs ne germait encore : car l' Éternel

Typologie
Logiciels

Avec KWUT : carte des sections

The screenshot shows the Lexico 3 interface with a search for 'de'. The 'Sections' window displays a KWUT map for the word 'de' in a biblical text, with sections highlighted in yellow:

de
 et
 la
 l'
 le
 les
 à
 il
 qui
 d
 des
 que
 est
 en
 Éternel
 dans
 vous
 Et
 un
 pour
 du
 ne
 je
 pas
 qu
 ils
 ce
 a
 lui
 tu
 Dieu
 sur
 se
 dit
 fils

The 'TextPloreur' window shows a KWUT analysis of the word 'de' in a biblical text, with sections highlighted in yellow:

de
 et
 la
 l'
 le
 les
 à
 il
 qui
 d
 des
 que
 est
 en
 Éternel
 dans
 vous
 Et
 un
 pour
 du
 ne
 je
 pas
 qu
 ils
 ce
 a
 lui
 tu
 Dieu
 sur
 se
 dit
 fils

Typologie
Logiciels

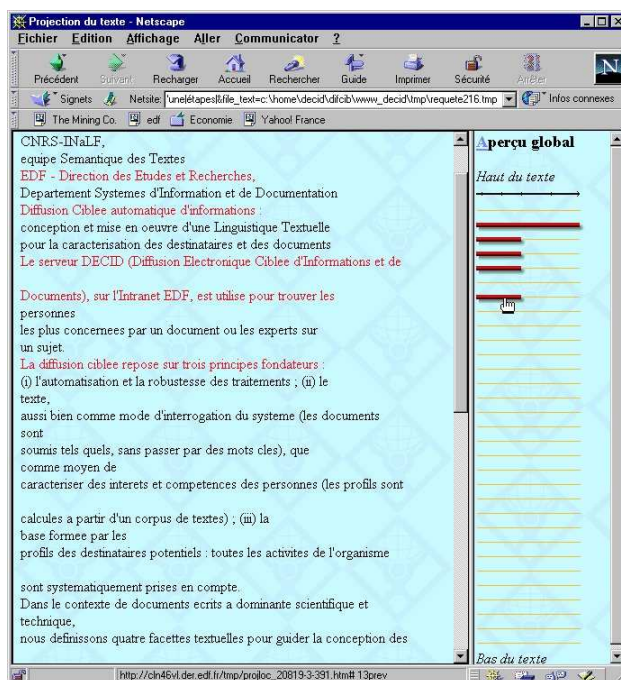
Histogramme marginal

(proposé dans un autre
contexte)

Fenêtre = vue locale

Histogramme = vue globale

Ascenseur = articulation
global / local

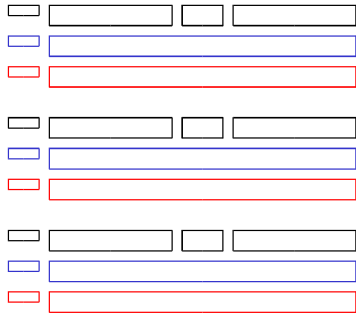


Exemple d'application : concordancier multilingue

- Corpus parallèle, aligné par exemple au niveau du paragraphe

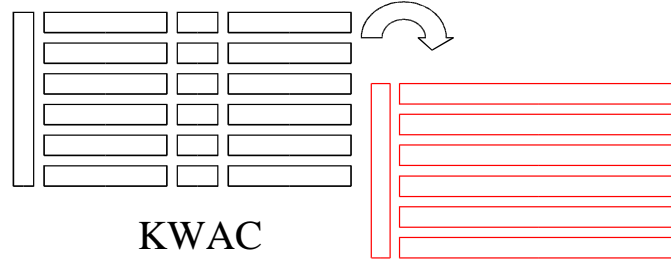
Des concordanciers multilingues non satisfaisants

1) Présentation intercalée :



Rupture des effets visuels d'alignement et répétition

2) Projection :



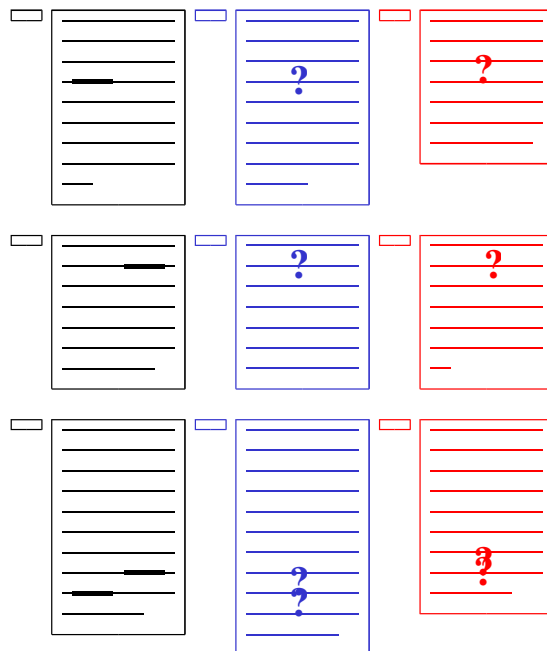
KWAC

Faux KWAC

(KWIC sur une ligne, appauvri)

Proposition pour KWIC multilingue

Colonnes étroites :
repérage visuel
début / milieu / fin



Proposition pour KWAC multilingue



+ concordances (KWAC) en langue cible

autonomes (générées dans la langue cible considérée) *et*

liées (possibilité par exemple de colorer une à une les lignes de contexte qui se correspondent, ou de les filtrer globalement)

KWIC traditionnel

Luhn (1958), pour la génération automatisée
d'index de titres de publications scientifiques

SdN06	Semaine du	Document	numérique
Albi06	Colloque	Documents	numériques et interprétation
Albi06	Colloque Documents numériques et	interprétation	
SdN06	Semaine du Document	numérique	
Albi06	Colloque Documents	Numériques	et interprétation

Mais :

- 2) Perte de place
- 3) Difficulté de mise en correspondance de la référence avec le titre

KWAC et KWOC traditionnels

SdN06	Document numérique Semaine du
Albi06	Documents numériques et interprétation Colloque
Albi06	interprétation Colloque Documents numériques et
SdN06	numérique Semaine du Document
Albi06	numériques et interprétation Colloque Documents

KWAC
KeyWord
Alongside
Context

SdN06	Document	Semaine du Document numérique
Albi06	Documents	Colloque Documents numériques et interprétation
Albi06	interprétation	Colloque Documents numériques et interprétation
SdN06	numérique	Semaine du Document numérique
Albi06	numériques	Colloque Documents numériques et interprétation

Du KWAC au KWOC : gain en lisibilité

KWOC

KWOC seconde forme (avec factorisation)

Document(s)

Semaine du Document numérique (*SdN06*)

Colloque Documents numériques et interprétation (*Albi06*)

Interprétation

Colloque Document numérique et interprétation (*Albi06*)

Numérique(s)

Semaine du Document numérique (*SdN06*)

Colloque Documents numériques et interprétation (*Albi06*)

- Cf. Index de fin de livre !
- Tri à l'intérieur de chaque rubrique =
Double-KWIC

Décalages terminologiques

Caractéristique	Existant	Proposé
Superposition et tri	KWIC	KWAC
(Rotation)	KWAC	-
Relevé de contextes	KWOC	KWIC
Relevé d'attestations	-	KWOC

... il faudrait trouver une autre série de dénominations ?...

Les concordances traditionnelles

- Apparemment plutôt KWIC car
 - Relevé de contextes
- Mais non sans lien avec les KWAC car
 - Contextes d'une ligne
 - Mise en valeur des parallélismes (syntaxico-sémantiques)

Spécificités

- Concordance manuelle
 - Stable et globale
 - Équilibre synthèse/détail
 - Pertinence de la délimitation des contextes
 - Une lecture qui fait autorité
- Concordancier
 - Dynamique et locale
 - Régularité et exhaustivité des relevés
 - Parallélismes visuels
 - Au lecteur de se donner des relevés pertinents

Même principe herméneutique fondamental : parallélismes et contrastes

Proposition technique d'amélioration des concordanciers

- Le pivot se détaille comme une séquence de **zones** successives
- Pour chaque zone possibilité de :
 - 1) Empilement en colonne
 - 2) Mise en valeur typographique (gras, couleur...)
 - 3) Tri (alphabétique, hiérarchique, canonique)

Zones : exemple

	<i>Contexte gauche</i>	shall	- MOT{0,3}	- be .+ed	+ <i>Contexte droit</i>
1	Pas de colonne	Pas de colonne	colonne	colonne	Pas de colonne
2	Normal	Normal	<i>Rouge + Italique</i>	Vert + Gras	Normal
3	Pas de tri	Pas de tri	2, Alphabétique	1, Fréquence	3, Alpha.

Zones : exemple de résultat

... Such declarations shall		be deposited	by the St...
... equally authentic , shall		be deposited	in the ar...
...
... Such gratis personnel shall		be employed	in accorda..
... under 18 years of age shall	<i>not</i>	be employed	in night w..
subject to compulsory education shall	<i>not</i>	be employed	in such wo.
...
... nor life imprisonment [...] shall		be imposed	for offence.
... was committed . Nor shall	<i>a heavier penalty</i>	be imposed	than the on
... was committed . Nor shall	<i>a heavier penalty</i>	be imposed	than the on
... Sentence of death shall	<i>not</i>	be imposed	for crimes

Sémantique des concordances

- Contextualisation dense
 - Contexte syntagmatique de chaque attestation
 - Contexte paradigmatique des entrées
 - Parallélismes synoptiques, accentués par superposition et tris
 - Contexte textuel et intertextuel (références)
- => cf. sémantique textuelle et différentielle (Rastier 2001)

Le « bon KWAC » ?

- Cultiver les atouts spécifiques et jouer la complémentarité :
 - Relevés d'attestations (KWOC, KWAC, KWIC, KWUT)
 - Concordance traditionnelle papier et concordancier sur édition numérique
- La pertinence herméneutique des parallèles contrastifs
 - Volet théorique : sémantique textuelle différentielle
 - Volet pratique : alignements, tris, zones