

1. Introduction

La concordance de la Banque de données textuelles de Sherbrooke (BDTS) est un outil d'analyse de la langue française qui permet de rechercher et de visualiser les occurrences d'un mot ou d'une séquence de mots dans un corpus textuel. Elle est basée sur la technologie de la linguistique computationnelle et est utilisée pour l'enseignement de la langue française, la recherche en linguistique et la documentation.

De plus, la BDTS est un outil d'analyse de la langue française qui permet de rechercher et de visualiser les occurrences d'un mot ou d'une séquence de mots dans un corpus textuel. Elle est basée sur la technologie de la linguistique computationnelle et est utilisée pour l'enseignement de la langue française, la recherche en linguistique et la documentation.

2. Banque de données textuelles de Sherbrooke (BDTS)

La BDTS est une banque de données textuelles de la langue française qui permet de rechercher et de visualiser les occurrences d'un mot ou d'une séquence de mots dans un corpus textuel. Elle est basée sur la technologie de la linguistique computationnelle et est utilisée pour l'enseignement de la langue française, la recherche en linguistique et la documentation.

¹, d

² par concordance

¹ Pour plus d'informations sur la BDTS, voir le site web de la Banque de données textuelles de Sherbrooke.

² De plus, la BDTS est un outil d'analyse de la langue française qui permet de rechercher et de visualiser les occurrences d'un mot ou d'une séquence de mots dans un corpus textuel.

Le sous-ensemble de langue spécialisée

3.

2.1. Typologie et composition de la BDTS

Les 37 BDTS sont composées de 59 % de langues (22 langues), 16 % de dialectes (6 dialectes), 14 % de langues régionales (6 langues), 6 % de langues étrangères (5 langues) et 5 % de langues créoles (2 langues).

4. La langue régionale

Le sous-ensemble de langue régionale est composé de 10 000 mots et est représenté par 1960 mots.

Le sous-ensemble de langue spécialisée est composé de 1206 000 mots et est représenté par 192 000 mots. Le sous-ensemble de langue littéraire est composé de 1206 000 mots et est représenté par 192 000 mots. Le sous-ensemble de langue journalistique est composé de 1206 000 mots et est représenté par 192 000 mots. Le sous-ensemble de langue didactique est composé de 1206 000 mots et est représenté par 192 000 mots. Le sous-ensemble de langue orale est composé de 1206 000 mots et est représenté par 192 000 mots.

Le sous-ensemble de langue spécialisée

Le sous-ensemble de langue spécialisée est composé de 1206 000 mots et est représenté par 192 000 mots.

Le sous-ensemble de langue littéraire

Le sous-ensemble de langue littéraire est composé de 1206 000 mots et est représenté par 192 000 mots.

Le sous-ensemble de langue journalistique

Le sous-ensemble de langue journalistique est composé de 1206 000 mots et est représenté par 192 000 mots. Les journaux (La Presse, Le Devoir, Le Soleil, Le Droit, L'Actualité, Voir, Québec Science, Franc-Vert, Interface, etc.)

Le sous-ensemble de langue didactique

Le sous-ensemble de langue didactique est composé de 1206 000 mots et est représenté par 192 000 mots.

Le sous-ensemble de langue orale

Le sous-ensemble de langue orale est composé de 1206 000 mots et est représenté par 192 000 mots. Les langues créoles (le français québécois, le français acadien, le français de la Gaspésie, etc.)

2.2. Brève caractérisation de la BDTS

La BDTS, sous-ensemble de langue spécialisée

3 Letm standard, sous-ensemble de langue spécialisée

4 Noms de lieux (Québec, Montréal, etc.)

à la Bas
 BDTS. Les
 é

<p> un songe éveillé. Comment ne pas se i des compensations l'incitait à se irait? Jamais plus il ne pourrait se chaleur de la maison acheva de les it du fond du seau finissait par me eine. L'inconnu ne pouvait que nous à l'autre, à l'ivre enfant Qui vous geaient, qui la soutenaient, qui la quelques gorgées de vin mousseux la P73 saveur aux aliments. Ce qui la en pointe, l'amiral, timidement, se et au fur_ et à mesure que le vin me is Chateaubriand», et cette idée me Il pourrait lui parler. Déjà il se ui peut se révéler particulièrement en va et, en même temps, c'est très mptueux que le Forum, ait été moins banques ont mieux résisté au climat </p>	<p> *griser d'illusions, comment ne pas renaît *griser des colères qu'il pouvait faire à *griser de_l espérance d'un bonheur venant *griser en battant aux tempes et contre le *griser. Puis le lait s'ajoutant au lait, *griser. C'étaient des images d'un audacie *grisa de mots, certain soir triomphant. C *grisaient du fond de la glace. Plus tard, *grisaient plus_qu un demi#litre de whisky *grisait surtout, c'était dans une glace p *grisait tous les soirs. Les deux P140 so *grisait. Cela devenait grossier. En elle *grisait. J'avais le sentiment de communie *grisait d'images et d'attouchements. Mais *grisant pour ceux qui apprennent à écrire *grisant parce qu'on est les premiers à le *grisant, c'est compréhensible. Que Joe Ca *grisant des marchés financiers d'alors. M </p>
--	--

Figure 1. Exemple du traitement du vocable « griser » et de ses formes à la première étape

A
 ay
 à
 de
 à
 co
 nq

A
 de
 p
 b
 XML
 p
 b
 XML
 f

L
 à
 D
 n
 n
 s
 g
 q
 n
 n
 C
 D
 C
 c

fiche CNC. Les

Entrée, (2) Cooc-

currents et constructions, (3) Remarque (4) Rédaction. Les

« *Entrée* »

Eltsbissit (dsb) bitté
 éam illa isipén

- dtré ch caignit adab (af) ité ds
 ÉT h d h (TD) ta f d sh BDS. Ilcté
 cse T h d h c p h s n n
 t p d h d h s
- d f p c t D a n c t p e s t s f o n
 t p h c o d s a e t t e
- d p t e b i (a f < g r > a n f i d a
 e h) S o c t e t s t e g t a c k a u t e
 f d s t h d h (d d s d h) t p
 d i n c h a s s e.

« *Cooccurents et constructions* »

C p t h É t c h d i t t h É l e p a b h u j
 d p s c o c t h d n p h e o c t a f e
 d t s p s c t h BDS.

« *Remarque* »

A u b a s h É h u p c e (d h l s p a c c e t h o r
 d l c t h BDS-c o d e s

« *Rédaction* »

T r a n s É t h d h t a s É h B a s t h
 g (B D I) , à É t c p e p a s h t a p h i n
 ó s a u t h B D I.

4. Un exemple de traitement : le verbe *atterrir*

4.1. Les modalités de la construction de la fiche CNC (BDTS-concordances)

À p p l a g d e p s e j n e d s h É h X M L
 c p h k e s h B a s T h s D o n c (T D) h é h
 É E r f e s s h o u t t h i p e j
 t p s t h f i c e j s a p t o s c h i s e n
 a i g b h d n s e d d h e i t p r c q b n
 s h a B D T S n y o s p I l a g l b d c b d s
 c a s c o c t a É h C N C , s e s t D , a f i d e (à s h d
 t a t t e c h h h i c h É h E n s c h p l a g
 d s h c a s c o s a s e à c h e p e c t h e c t e
 (É h B D T S) É t p a d h h t h c a

4.2. Le résultat du traitement et la fiche CNC ainsi constituée

U n c h a É h q u i t e h f o t t h (1 5 6) , c h s l
 b i t s h u t h h i t s c o c t (2 5 9) , i d
 g h e g r e b e s s e e . , s p a n h
 a t t e r r i t h (X S L) d s h c i

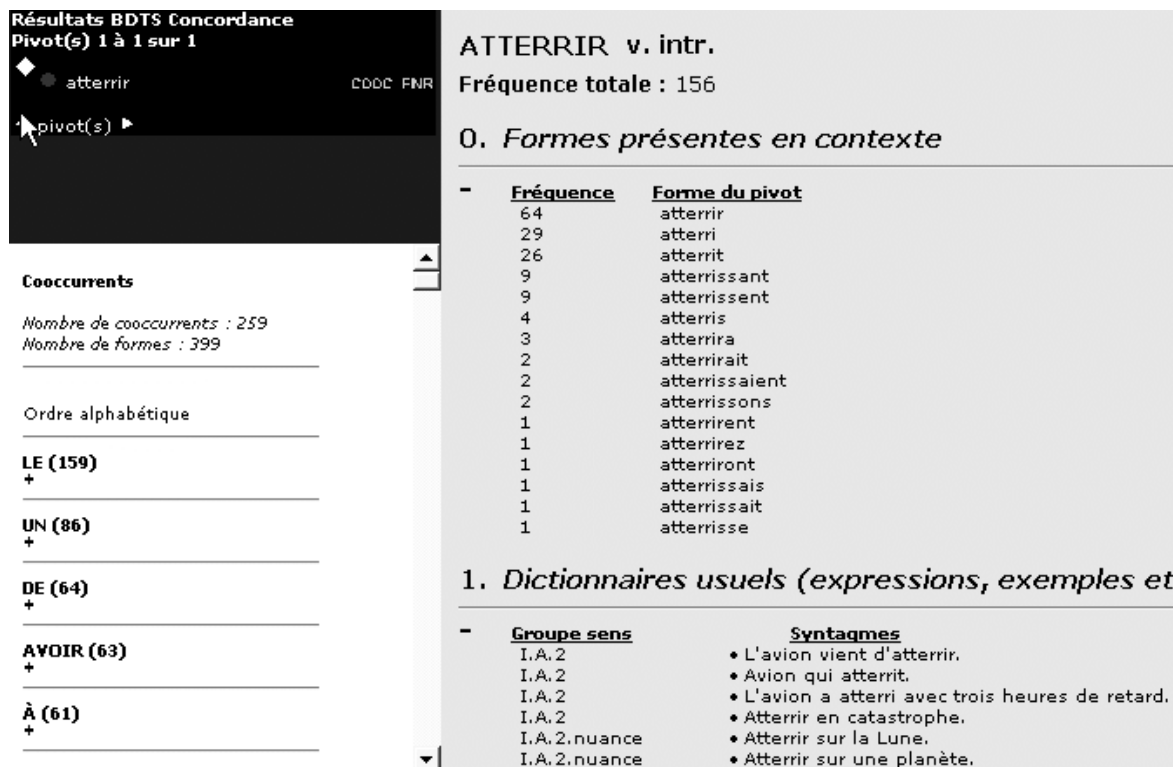


Figure 2. Vue très fragmentaire de la fiche CNC (BDTS-concordances) du verbe « atterrir »

5. Intérêt et limites de la BDTs-concordances

La CNC a été développée par le Centre de Recherches Linguistiques sur les Langues et les Langues (CRL) de l'Université de Caen. Elle est basée sur le corpus de la BDTs (BDTS-concordances) et est accessible en ligne.

5.1. L'intérêt des fiches CNC (BDTS-concordances)

Les fiches CNC (BDTS-concordances) sont basées sur le corpus de la BDTs (BDTS-concordances) et sont accessibles en ligne. Elles fournissent des informations sur les formes grammaticales et les syntagmes du verbe.

À la CNC/XML, les fiches CNC (BDTS-concordances) sont basées sur le corpus de la BDTs (BDTS-concordances) et sont accessibles en ligne. Elles fournissent des informations sur les formes grammaticales et les syntagmes du verbe.

avant!

Il est un fait que tous les aspects (à l'exception de la fréquence) sont pris en compte dans la classification. La classification est basée sur la fréquence de l'usage de ce mot dans les différents contextes.

<i>GRISER [fréquence = 52 après tri des contextes]</i>			
###[(<i>dp</i>) GRISER QQN : A] [<i>vnsd</i>]		[5]	
###[(<i>dp</i>) SE GRISER : S] [<i>vpfh</i>]		[3]	
###[(<i>Édp</i>) GRISER QQN DE : A] [<i>vnsd</i>]			[17]
###[(<i>Édp</i>) ÊTRE GRISÉ, PAR : S] [<i>ipf</i>]			[11]
###[(<i>Édp</i>) GRISÉ S] [<i>pt psaj</i>]		[2]	
###[(<i>Édp</i>) SE GRISER DE QQCH : A] [<i>vpfh</i>]			[13]
###[(<i>n</i>) GRISER QQCH : S] [<i>vnsd</i>]	[1]		
###[IGNORÉ]	[2]		

Tableau 1. Exemple du contenu du dossier-sens pour le vocable « griser »

Il est un fait que tous les aspects (à l'exception de la fréquence) sont pris en compte dans la classification. La classification est basée sur la fréquence de l'usage de ce mot dans les différents contextes.

- P. (2000). *Dictionnaires et nouvelles technologies*. Paris: [L'Asie](#).
 T. A. & W. C. (1998). *Informatisation du Dictionnaire universel de Furetière revu par Basnage de Bauval (1702) : premier bilan*. [DOI: 10.1017/S0013716519980001](#).
 U. & V. (1996). *TEI Guidelines for Electronic Text Encoding and Interchange*. [DOI: 10.1017/S0013716519960001](#).
 V. & N. (1996). *Encyclopédie de la philologie*. Paris: [L'Asie](#).
 T. E. I. (Ed), *Lexicographie et informatique. Autour de l'informatisation du Trésor de la langue française. Actes du Colloque international de Nancy*. [DOI: 10.1017/S0013716519960001](#).
 W. R. (1997). *Baliser un texte, c'est le penser : le cas du Dictionnaire de l'Académie française*. [DOI: 10.1017/S0013716519970001](#).