

Collecting, Encoding and Organizing Collocates in a Terminological Database

Marie-Claude L'Homme

Observatoire de linguistique Sens-Texte (OLST)

Université de Montréal

mc.lhomme@umontreal.ca

www.olst.umontreal.ca

Outline

- The DiColInfo: a lexical resource for Terminology
 - Some articles
 - Brief description of the methodology

- Collocations in terminology
 - Previous work

- Collocations in the DiColInfo
 - Definition of “collocation”
 - Collection of collocations
 - Encoding of collocations
 - Ordering of collocations

The DiCoInfo: a lexical resource for terminology

- A French terminological database with terms in the fields of computing and the Internet
- A description of lexico-semantic properties of terms
 - Semantic distinctions
 - Actantial structure
 - List of paradigmatic and syntagmatic relations between terms
- Terms belong to different parts of speech
 - nouns (including noun phrases) : *courriel* (Eng. email)
 - verbs: *configurer* (Eng. configure); *clavarder* (Eng. chat)
 - adjectives and adverbs: *convivial* (Eng. user-friendly)

Some articles (1)

■ Current version

- <http://olst.ling.umontreal.ca/dicoinfo/>
- Descriptions written by a team of terminologists (research group Observatoire de linguistique Sens-Texte, OLST)
- Formal representations developed in collaboration with a team from the research group Recherche appliquée en linguistique informatique (RALI, University of Montreal) : Guy Lapalme, Benoît Alain et Vincent St-Amour
- XML representation (result of a conversion of an Access database)
- Edited directly in the XML database
- Website generated from the XML representation

Some articles (2)

[Aide](#)

[Liste alphabétique des termes](#)

Rechercher

Résultats de la recherche : 6 lexie(s)

blogage [1](#)
blogosphère [1](#) [2](#)
blogue [1](#)
bloguer [1](#)
blogueur [1](#)

blogage [1](#), n. m.

Statut : 1

Structure actancielle : du blogage : ~ de AGENT-A (internaute 1) avec AGENT-B (internaute 1)

[Réalizations linguistiques des actants](#)

[Contextes](#)

[Liens lexicaux](#)

Rédacteur(s) : MCLH

Date de mise à jour : 8/7/2007

Some articles (3)

Head word

Résultats de la recherche : 1 lexie(s)

partitionner 1

Actantial structure

partitionner 1, v. tr.

Statut : 0

Structure actancielle : partitionner : AGENT{utilisateur 1} ~ PATIENT{disque 1}

Réalisations linguistiques des actants

Définition : Un UTILISATEUR subdivise un DISQUE en parties souvent représentées comme des disques séparés.

Contextes

Liens lexicaux

Rédacteur(s) : MCLH

Date de mise à jour : 5/8/2007

Definition

Some articles (4)

Résultats de la recherche : 1 lexie(s)

partitionner 1

partitionner 1, v. tr.


Statut : 0

Structure actancielle : partitionner : AGENT{utilisateur 1} ~ PATIENT{disque 1}

Réalisations linguistiques des actants

agent
utilisateur 1
patient
disque 1 , disque dur 1

Linguistic realizations of actants



Définition : Un UTILISATEUR subdivise un DISQUE en parties souvent représentées comme des disques séparés.

[Contextes](#)

[Liens lexicaux](#)

Rédacteur(s) : MCLH

Date de mise à jour : 5/8/2007

Some articles (5)

Résultats de la recherche : 1 lexie(s)

[partitionner](#) 1

partitionner **1**, v. tr.

Statut : 0

Structure actancielle : partitionner : AGENT(utilisateur 1) ~ PATIENT(disque 1)

Réalisations linguistiques des actants

Définition : Un UTILISATEUR subdivise un DISQUE en parties souvent représentées comme des disques séparés.

Contextes

Vous devrez donc affecter cet identificateur à votre partition de swap lorsque vous partitionnerez votre disque dur. (Source : LINUX3P1) /

Partitionnement du disque dur: cette distribution vous permet de partitionner votre disque à ce niveau de l'installation (si vous ne l'avez pas fait précédemment) et vous laisse le choix entre disk druid et fdisk (Source : LININT) / En effet, partitionner un disque suppose qu'il existe déjà des pistes et des secteurs sur le disque. (Source : LINUX3P1)

Liens lexicaux

Rédacteur(s) : MCLH

Date de mise à jour : 5/8/2007

Some articles (5)

^{ks}
partitionner 1, v. tr.

Statut : 0

Structure actancielle : partitionner : AGENT(utilisateur 1) ~ PATIENT(disque 1)

[Réalizations linguistiques des actants](#)

Définition : Un UTILISATEUR subdivise un DISQUE en parties souvent représentées comme des disques séparés.

[Contextes](#)

[Liens lexicaux](#)

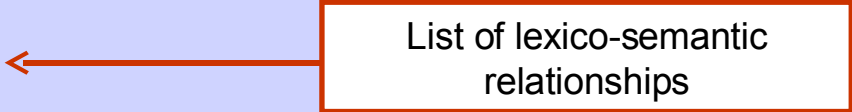
[Rôles thématiques](#)

[Fonctions lexicales](#)

(Ces informations ne sont disponibles que sous Firefox)

Explication - terme typique	Lexie reliée
≈	défragmenter 1
≈	formater 1
≈	initialiser 1
≈	reformatier 1
Nom	partitionnement 1
Résultat	partition 1
Un disque qui a été p.	partitionné 1

List of lexico-semantic relationships



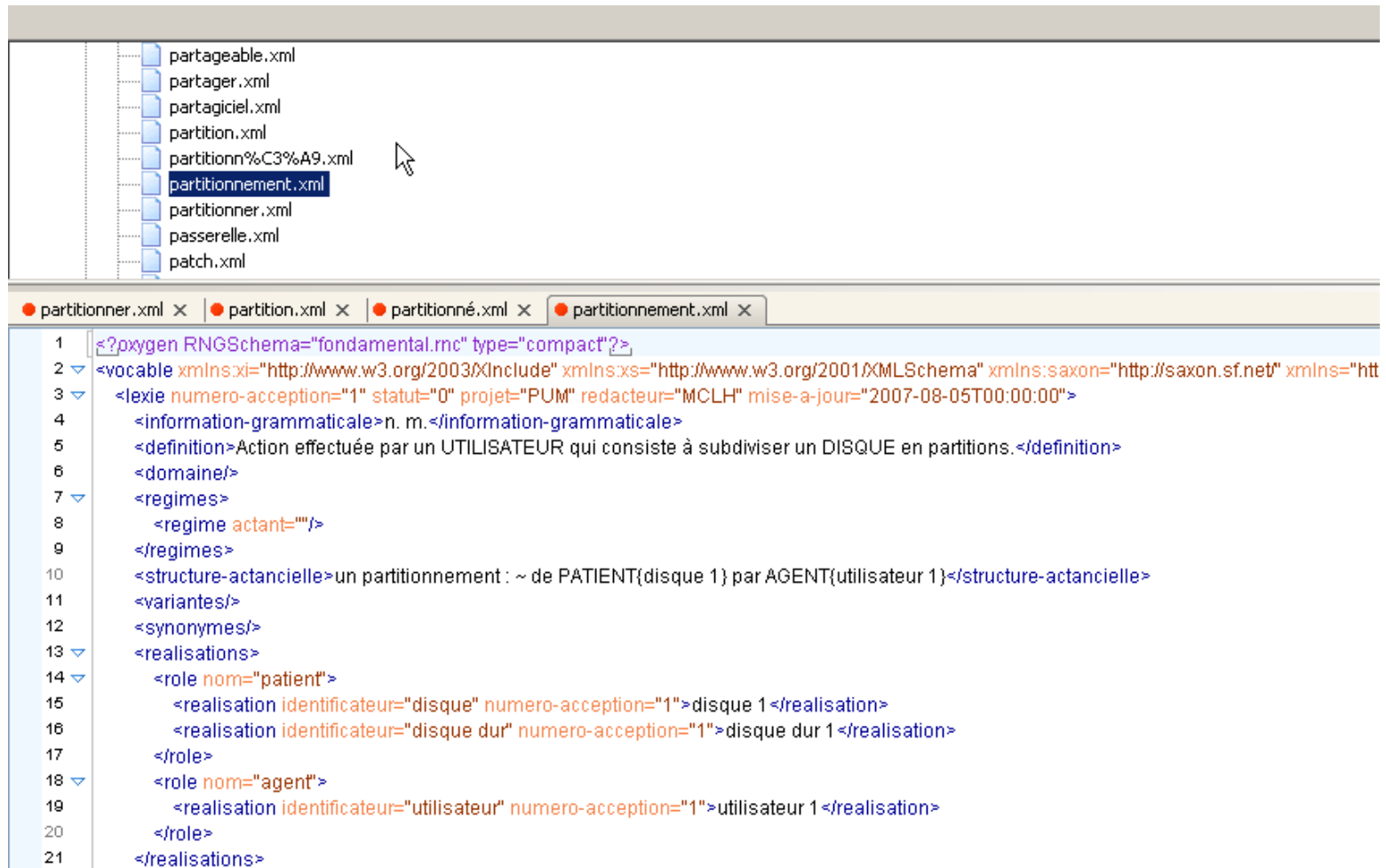
Rédacteur(s) : MCLH

Date de mise à jour : 5/8/2007

Methodology

- Use of a corpus
 - Approx. 1 million words
- Selection of terms
 - Term extraction (TermoStat, Drouin 2003)
 - Application of lexico-semantic criteria
- Semantic distinctions
 - Application of lexico-semantic criteria
- Definition of the actancial structure and list of actants
- List of related terms
- Writing of the definition

Editing in <oxygen/>



The screenshot displays the Oxygen XML Editor interface. The top-left pane shows a file explorer with a list of XML files: partageable.xml, partager.xml, partagiciel.xml, partition.xml, partitionn%C3%A9.xml, **partitionnement.xml** (selected), partitionner.xml, passerelle.xml, and patch.xml. The top-right pane shows the editor tabs: partitionner.xml, partition.xml, partitionné.xml, and partitionnement.xml. The main editor window displays the XML content for partitionnement.xml, with line numbers 1 through 21 on the left. The XML code is as follows:

```
1 <?oxygen RNGSchema="fondamental.mc" type="compact"?>
2 <vocabulary xmlns:xi="http://www.w3.org/2003/XInclude" xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns:saxon="http://saxon.sf.net" xmlns="ht
3 <lexie numero-acception="1" statut="0" projet="PUM" redacteur="MCLH" mise-a-jour="2007-08-05T00:00:00">
4   <information-grammaticale>n. m.</information-grammaticale>
5   <definition>Action effectuée par un UTILISATEUR qui consiste à subdiviser un DISQUE en partitions.</definition>
6   <domaine/>
7   <regimes>
8     <regime actant=""/>
9   </regimes>
10  <structure-actancielle>un partitionnement : ~ de PATIENT{disque 1} par AGENT{utilisateur 1}</structure-actancielle>
11  <variantes/>
12  <synonymes/>
13  <realisations>
14    <role nom="patient">
15      <realisation identificateur="disque" numero-acception="1">disque 1</realisation>
16      <realisation identificateur="disque dur" numero-acception="1">disque dur 1</realisation>
17    </role>
18    <role nom="agent">
19      <realisation identificateur="utilisateur" numero-acception="1">utilisateur 1</realisation>
20    </role>
21  </realisations>
```

Current state of the DiCoInfo

Articles on line (Status 0)	approx. 226
Articles on line (Status 1)	approx. 703
Articles on line (Status 2)	approx. 37
In preparation	approx. 900
Total number of lexical relations described	Approx. 28 780
Total number of lexical relations on line	approx. 22 300
Ratio lexical relations / completed articles	approx. 23
Korean version	approx. 100 articles
English version	approx. 600 articles
Spanish version	Terms being selected

Collocations in terminology (1)

- Cohen (1986)
 - Collocations in the field of stock options (French)
 - Organization based on the parts of speech of collocates and their general meaning
 - *Profit* : (subject + verb “increase”) : ~ *s’accroît*, ~ *s’améliore*, ~ *augmente*, ~ *monte*, etc.
- Meynard (2000)
 - Collocations in the field of the Internet (bilingual English – French)
 - Organization based on the parts of speech of collocates
 - *Address* (collocate verb + noun) : *activate an ~*, *assign an ~*, *attach an ~*, *copy an ~*, *delete an ~*, etc.

Collocations in terminology (2)

■ Termium

- Collocations on some term records (French)
- Organization based on the parts of speech of collocates
 - *Logiciel* (PHR – verbs) : *construire, implémenter, interfacer, réaliser un logiciel, simuler, etc.*

■ Dictionnaire d'apprentissage du français des affaires (DAFA), Binon et al. (2000)

- Collocations in the field of business (multilingual dictionary, but collocates are provided for French terms)
- Organization based on the parts of speech of collocates and their general meaning
- Description of the actancial structure of the key word

Collocations in terminology (3)

■ DAFA

■ concurrence (Engl. competition)

- **+ adjectif** (“type de concurrence”) *la libre concurrence* (explanation); *une concurrence directe* (explanation); *une concurrence monopolistique* (explanation), etc.
- **+ nom** : *la pression de la concurrence sur X* (explanation); *la concurrence du grand nombre* (explanation), etc.
- **+ verbe**, qui fait quoi ? : *X entre en ~ avec Y*, *X est en ~ avec Y*, *X et Y se livrent une ~*, etc.

Definition of collocation (1)

- Partly based on Explanatory Combinatorial Lexicology, ECL (Mel'čuk et al. 1995)
 - Unpredictable combination of lexical units
 - A base selected by the speaker to express a given meaning
 - A collocate the selection of which is imposed by the base
 - In our case, the base is defined as a term
 - “typical use”
 - *Internet: surf the Internet*
 - *file: edit a file*
 - *mouse: move a mouse*
 - *password: use a password*

Collection of collocations

- Analysis of concordances extracted from the 1 million word corpus
 - Nouns: verbal, nominal and adjective collocates
 - Adjectives: adverbial collocates (few)
 - Verbs: adverbial collocates (few)
- Analysis of the corpus partly automated
 - Claveau and L'Homme (2006)
- Queries on the Web to find other relevant collocates
- Frequency is taken into account

Encoding of collocations (1)

- Aspects to take into account when describing a collocation (Mel'čuk et al. 1995)
 - The syntactic relationship between the base and the collocate
 - *Move a mouse*: verb + 1st complement
 - *Click on something with a mouse*: click + 2nd complement
 - The actantial structure of the base
 - Mouse: used by someone (Act1) to act on something (Act2)
 - Move + mouse: Act1 uses the mouse
 - Click + mouse: Act1 uses the mouse to act on Act2
 - The sense of the collocate when combined with the base
 - Move + mouse: “typical use of the mouse”
 - Create + password: “creation of the password”

Encoding of collocations (2)

Internet

Internet: ~ utilisé par AGENT{utilisateur 1} pour intervenir sur
PATIENT{ressource 2, site 1}

L'utilisateur utilise I.

naviguer dans ~

L'utilisateur utilise I. pour intervenir sur
la ressource ou le site

chercher ... dans ~

trouver ... dans ~

Encoding of collocations (3)

barre d'espacement 1, n. f.

Statut : 0

Structure actancielle : la barre d'espacement : ~ utilisée par AGENT(utilisateur 1) pour intervenir sur PATIENT(espace 2)

[Réalisations linguistiques des actants](#)

Définition : Touche de forme allongée placée au bas du clavier qu'un UTILISATEUR manipule afin d'insérer des ESPACES.

Synonyme(s) : barre espace, barre d'espace

[Contextes](#)

[Liens lexicaux](#)

[Rôles thématiques](#)

[Fonctions lexicales](#)

(les deux liens ci-dessus sont fonctionnels avec le navigateur Firefox)

Explication - terme typique	Explication - rôle thématique	Fonction lexicale	Lexie reliée
≈	Générique	Gener	touche 1
≈	Sens voisin	Cf	pavé 1
≈	Sens voisin	Cf	flèche 1
L'utilisateur utilise la b.	L'agent utilise le « mot clé »	Real1	enfonceur 1 la ~
L'utilisateur utilise la b.	L'agent utilise le « mot clé »	Real1	appuyer 1 sur la ~
-> NOM	-> NOM	S0Real1	enfonceur 1 de la ~
-> NOM	-> NOM	S0Real1	appui 1 sur la ~
L'utilisateur utilise la b. pour intervenir sur un espace	L'agent utilise le « mot clé » pour intervenir sur le patient	Labreal12	insérer 2 ... avec la ~
-> NOM	-> NOM	S0Labreal12	insertion 2 de ... avec la ~
L'utilisateur cesse d'utiliser la b.	L'agent cesse d'utiliser le « mot clé »	FinReal1	relâcher 1 la ~
-> NOM	-> NOM	S0FinReal1	relâchement 1 de la ~
Lieu où on trouve une b.	Lieu où on trouve un « mot clé »	Slac	clavier 1

Ordering of collocations (1)

■ mot de passe

- Types of ...
 - Que l'utilisateur remplace par un autre = ancien ~
 - Qui fonctionne bien = bon ~, ~ valide
 - Qui ne fonctionne pas bien = mauvais ~, ~ invalide
- By the user (different steps)
 - L'utilisateur crée un m. = créer, choisir, définir un ~
 - L'utilisateur a un m. = avoir un ~, posséder un m.
 - L'utilisateur n'a pas un m. = oublier son m., perdre son ~
 - L'utilisateur a un m. de nouveau = retrouver son ~
 - L'utilisateur fait en sorte qu'un m. puisse être utilisé = activer un ~

Ordering of collocations (2)

■ mot de passe

■ Actions by the base

- Le m. intervient sur le compte ou le site = le ~ protège ...
- Le m. cesse de fonctionner = le ~ expire, arrive à échéance

■ Use by another actant or external participant

- L'ordinateur ou le fournisseur se prépare à utiliser un m. = demander un ~
- L'ordinateur ou le fournisseur commence à utiliser un m. = vérifier un ~
- L'ordinateur ou le fournisseur utilise un m. = accepter un ~

Future work

- Finish the dictionary!
- Define a method for collecting collocates according the the part of speech and the sense of the base
- Automate the collection of collocates
- Provide a formal description of the syntactic structure of collocations
- Use in language for special purposes teaching?
- Use in natural language processing (NLP)?