Presenting collocates in a dictionary of computing and the Internet according to user needs

Anne-Laure Jousse (1), Marie-Claude L'Homme (1),
Patrick Leroyer (2) and Benoît Robichaud (1)

(1) Observatoire de linguistique Sens-Texte (OLST)
Université de Montréal
anne-laure.jousse|mc.lhomme|benoit.robichaud@umontreal.ca

(2) Center for lexicography
Aarhus University
pl@asb.dk

Abstract

This paper presents a novel method for organizing and presenting collocations in a specialized dictionary of computing and the Internet. This work is undertaken in order to meet a specific user need, i.e. that of searching for a collocate (or a short list of collocates) that expresses a specific meaning in a text production situation. The model we suggest is based on lexical functions (=LFs) that formally encode syntactic, argumental and semantic properties of collocations. LFs are grouped in larger semantic classes (e.g., USE, CREATE, PLACE SOMEWHERE, etc.). The model is in the process of being implemented in the online version of the dictionary. Users are prompted with generic meanings associated with the classes we created and they can then select verb and noun collocates that express more specific meanings. The article describes the model for grouping collocations and its implementation. Finally we present a small pilot study that was conducted in order to gather some user feedback on the usefulness of the method.

Keywords

Collocations, lexical functions, specialized dictionary, text production, computing, Internet.

1 Providing onomasiological access to collocates

An increasing number of dictionaries (general and specialized) present collocations within their entries. Some of these dictionaries are online (DAFLES, DiCE, DiCoInfo) and thus can provide different access paths to users based on how collocations were encoded in the first place. Depending on the dictionary use situation, users will generally consider the presentation of collocations as extremely useful information. In specialized language, collocations are essential to the production of specialized discourse in accordance with the writing and genre conventions used by professionals. Even if compilers of dictionaries may take it for granted that the issue of selecting the collocations is settled, they still face a number of challenges, among which are the following:

1. How should collocations be presented in specific entries, especially in online dictionaries in which high numbers of collocations are listed?

2. Should a description of their meaning accompany collocations? What should this description look like?

3. What should be done to ensure the functional usability of data presentation and access?

This article will attempt to answer all three questions focusing on a specific user need, namely that of a user searching for a collocate (or a short list of collocates) that expresses a specific meaning in a text production situation. In other words, the user knows which meaning should be expressed but does not know the specific, conventionalized wording itself. More specifically, our aim is to design an access method that would allow users consulting an online dictionary of computing and the Internet to obtain answers to the questions such as those listed below:

- Which verbs express the idea of "using" a dialog box (utiliser une boîte de dialogue)?[1] Answer: *activer, afficher, ouvrir une boîte de dialogue* (enable, display, open a dialog box)
- Which verbs express the typical activities carried out by a programmer? Answer: *le programmeur écrit …, débogue .., développe …, corrige, programme* (the programmer writes …, debugs …, develops …, programs)

The remainder of the article is organized as follows. Section 2 presents previous attempts at organizing and presenting collocations in dictionaries. Section 3 describes the DiCoInfo and gives more specific details about how collocations are encoded in this dictionary. Section 4 explains how collocations were grouped and how the new access method was implemented. Finally, Section 5 gives the details of a pilot study that was carried out in order to check if the model suited the targeted user needs. Section 6 briefly outlines future work.

2 Previous work

In recent years, a few (printed or electronic) dictionaries and lexical databases have attempted to present collocations according to one or several organizing principles. In printed dictionaries (cf. Tutin, 2010) or lexical databases, two main organizing methods are preferred. Most reference works organize collocations syntactically (LTP Dictionary of Selected Collocations [Hill & Lewis, 2002], Antidote [Charest *et al*., 2007]); others add a semantic layer to the syntactic classification that is seldom made explicit and that merely presents itself as lists of synonymous collocates (Le Robert [Le Fur, 2007], BBI Dictionary of English Word Combinations [Benson *et al*., 1997], Oxford Collocations Dictionary for Students of English [McIntosh, C., 2009]).

More sophisticated repositories combine several organization principles (semantic, morphological, and syntactic) (Base Lexicale du Français [Verlinde et al., 2006]), le *Dictionnaire d'apprentissage du français des affaires* ([Binon et al., 2009]), Elexiko [Klosa et al., 2006]). However, when a proper modeling of semantic relationships between lexical units and collocates is lacking, a complete automatic grouping of collocates in semantic categories can simply not be considered. This leads lexicographers to classify relationships in an ad hoc manner and often to resort to introspective methods (Cinkova and Hanks, 2010).

If an overall organization of collocations is considered from the point of view of the entire lexicon of a language, it can only be carried out based on a solid formalization of lexical

[1] The implementation of the method is currently carried out in the French version of the dictionary. However, an extension to the English and Spanish versions will be done shortly.

relationships. Indeed, this formalization allows for a generalization of the classification to larger parts of the lexicon (based on the development of a model on a representative sample): the vocabulary associated with a specialized subject field, the lexicon of a specific language or a group of languages. Formalization is also necessary for processing and manipulating data.

Hence, the formal system of Lexical Functions (=LFs) (presented in Section 3.3) used in Explanatory Combinatorial Lexicology (=ECL) is perfectly adapted for this kind of task. Various lexical resources based on this framework have been developed during the past few decades. Some take the form of lexical databases (DiCo [Polguère, 2000], DiCE [Alonso Ramos, 2004], DiCoInfo [L'Homme, 2011]); others are tools for language learning (Callex [Diachenko, 2006], Callex-Esp [Boguslavsky et al., 2006]). In all these, LFs are used to represent syntactic and semantic properties of lexical relationships.

A more specific proposal was made for representing relationships in a lexical database (Jousse, 2007; Jousse, 2010; Jousse et al. 2008). The authors use the existing database DiCo and suggest that paradigmatic as well as syntagmatic relationships be organized in such as way that users can follow different paths for browsing parts of the lexicon. The method takes into account: 1. a semantic organization that allows users to access relationships onomasiologically; 2. a syntactic organization for selecting a collocation based on a specific syntactic configuration; 3. a classification based on parts of speech; and, finally, 4. an organization taking into account communicative criteria that will suggest collocates according to the argument that is highlighted in a specific collocation (for example, in *X gives a call to Y*, X is highlighted, whereas Y is emphasized in *Y receive a call from X*). This method, however, has not been implemented yet.

The DiCoInfo, presented in the next section, is based in part on the same theoretical principles as the DiCo and thus lends itself to a similar formal organization of its collocations.

3 The DiCoInfo: form and functions

3.1 The DiCoInfo

The DiCoInfo, *Dictionnaire fondamental de l'informatique et de l'Internet* is an online dictionary (http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi) that provides information on terms pertaining the fields of computing and the Internet (e.g., *access, configure, dynamic, read, software*). Currently, the DiCoInfo contains over 1,000 articles in French and approx. 700 articles in English (a Spanish version is also under development).

The methodology for compiling the DiCoInfo is based on a combination of automated and manual methods. A series of steps (selection of terms, collection of example sentences, writing of entries) is carried out by terminologists more or less in the same order. The data is encoded in an XML structure and then converted into HTML pages for the purpose of publishing its content on the Internet.

3.2 Recent improvements for functional purposes

Work on a more adaptive and user-oriented access to data in the DiCoInfo was initiated back in 2009. It paved the way for the development of automatic access to translations of collocations (L'Homme & Leroyer, 2009). The software application was adapted a year later, and now includes a new version of the search engine, enabling user-friendly, automatic access to translations of collocations in French, English, and Spanish. Collocations sharing identical

semantic and syntactic properties were linked up by means of the encoding of LFs (L'Homme, Leroyer & Robichaud, 2010).

Providing onomasiological access to collocations was also considered. This could be done in part by using the paraphrasing of lexical relations in the database (cf. Section 3.3). The idea was to provide assistance in text production situations (in L1 or in L2) in which the user knows the meaning of a phraseological unit but is searching for the appropriate collocates that appear in combination with that unit. However, at the time, this new type of access was devised but not concretely designed or implemented.

3.3 Lexical functions in the DiCoInfo

One of the most important data categories in the DiCoInfo is that of lexical relationships. Each entry contains a list of lexical units sharing with the head word a paradigmatic or a syntagmatic relationship (synonymy, antonymy, syntactic derivation, collocates, etc.). All lexically-related units are explained using two different systems: 1. the system of LFs (Mel'čuk et al., 1995, Mel'čuk et al., 1984-1999); 2. a less formal and language-dependent explanation designed to be more transparent for users (these explanations are based on a proposal made by Polguère, 2003). Table 1 gives a few examples of how the collocations are explained in the database.[2]

Key word	Collocate	Lexical function	Explanation
programmer	the ~ write …	Fact2	The p. acts on the program
dialog box	open a ~	Real1	The user uses a d.
program	quit a ~	FinReal1	The user stops using a p.
Internet	browse the ~	Real1	The user uses the I.
keyboard	enter … on a ~	Labreal12	The user uses a k. to act on the data
account	access an ~	IncepReal1	The user starts using an a.

Table 1: Collocations, lexical functions and explanations

In the online version of the dictionary, lexical relationships (among which collocations) were all listed in the form of a table. Collocations were presented in a section called "Combinations" that was very long and difficult to read in some entries. For example, in the article devoted to "fichier" (file), approximately 100 collocates were listed.

4 A model for grouping and browsing collocations

4.1 Grouping collocations in transparent classes

In the DiCoInfo, LFs were first grouped into more general semantic classes to allow users to access collocations onomasiologically (from the meaning to the collocate). We analyzed the relationships that had been encoded in the DiCoInfo and found that specific classes were dominant (for example, since the field of computing needs terms that denote entities, many collocates express the idea of USE or MAKE STH WORK). The semantic classes were defined based on the results of the analysis of corpus data in order to ensure that they would capture recurrent relationships in a balanced way. When defining the classes, terminologists started using some frequent collocational relationships. As was said above, LFs encoding USE and TO MAKE STH WORK ($Real_i$, $Labreal_{ij}$), etc. were particularly productive. All LFs

[2] Wherever possible, each LF is explained with a unique gloss. As far as the phrasing of glosses is concerned, we aim to provide – as much as possible – a transparent and natural explanation. In addition, the phrasing may vary slightly according to the base of the collocation.

encoding a typical use were first grouped regardless of arguments and secondary meanings involved ("Use", "Use for something", "Agent uses", "Other argument or external participant uses") into intermediate classes. Then, more generic classed were defined. While conducting this analysis, we also took into consideration that the user might need to access specific pieces of information concerning collocates.

Lexical function	Intermediate class	Generic class
Caus1Able1Fact0, Caus1Able1Real1, Caus1Able1Real3, PermFact0, Perm1Fact0	**Permettre l'utilisation / Activer**	**UTILISER / NE PAS UTILISER**
Prepar@, Prepar1, Prepar1Fact0, Prepar1Real1, Prepar2Real3, Prepar@Fact0, De_nouveauPrepar1, De_nouveauPrepar1Fact0	**Préparer l'utilisation / le fonctionnement**	
IncepLabreal12, IncepReal@, IncepReal1, IncepReal2, IncepReal3	**Commencer à utiliser / Apparaître**	
Caus1Fact0, Caus@Fact0, Labreal@2, Labreal12, Labreal123, QLabreal12, Real@, Real1, Real12, Real123	**Utiliser / Faire fonctionner**	
FinLabreal12, FinReal1, FinReal2, Liqu1Fact0, Liqu@Fact0	**Cesser d'utiliser / de faire fonctionner**	

Table 2: Grouping of LFs under intermediate and generic classes

CRÉER/SUPPRIMER **Créer / Faire apparaître** *créer, générer un fichier*	**TO CREATE/TO DELETE** **To create or display** *to create, to generate a file*
Supprimer / Détruire *supprimer, effacer un fichier*	**To delete / eliminate** *to delete a file*
TRANSFORMER **Transformer** *crypter, convertir un fichier*	**TO TRANSFORM** **To transform** *to encrypt, to convert a file*
Diminuer / Réduire *comprimer un fichier*	**To reduce** *to compress a file*
UTILISER/NE PAS UTILISER **Préparer l'utilisation / Le fonctionnement** *installer, rechercher un fichier*	**TO USE/ USE NOT** **To prepare for use, operation** *to install a file, to search for a file*
Commencer à utiliser / Apparaître *charger, ouvrir un fichier*	**To start to use / to appear** *lo load, to open a file*
Utiliser / Faire fonctionner *traiter, éditer un fichier*	**To use/to make sth work** *to process, to edit a file*
Cesser d'utiliser / De faire fonctionner *fermer un fichier*	**To stop using/working** *to close a file*
METTRE QUELQUE PART **Ajouter à / Mettre dans** *joindre un fichier à un courriel*	**TO PLACE SOMEWHERE** **To add, to place in** *to attach a file to an e-mail*
Stocker ~ quelque part *archiver, télécharger un fichier*	**To store somewhere** *to archive, to download a file*
Transférer *exporter, transférer un fichier*	**To transfer** *to export, to forward a file*
Extraire / Sortir de *désarchiver un fichier*	**To extract/Quit** *to extract a file*
IDENTIFIER **Identifier** *nommer un fichier*	**TO IDENTIFY** **To identify** *to name a file*

Table 3: Classification examples for a subset of collocations of 'fichier'

Classes were defined according to the main relationships that can be observed in the field of computing and not according to general principles that could apply to general language (as in Jousse, 2010). Table 2 shows how we grouped recurrent LFs in intermediate classes under the generic class of **UTILISER/NE PAS UTILISER** (TO USE/USE NOT).

In fact, some of the classes identified could apply to other subject fields and to general language, but we focused on what could be observed in our database. We also believe that our method for grouping and balancing them is closely related to the subject field we are dealing with. Hence, it is most likely that some classes might not have the same value in other specialized subject fields. Examples in point are **METTRE QUELQUE PART** (TO PLACE SOMEWHERE, used to capture collocates such as *store* and *save*) and **TRANSFORMER** (TO TRANSFORM, used to capture verbs like *format* and *compile*).

We also tried to limit the number of different classes as much as possible in order not to overload the interface with long lists of classes and help users memorize them more easily. Up to now, 9 generic classes and 45 intermediate classes have been defined. Intermediate classes contain approx. 300 different LFs. Table 3 gives examples of classes that are displaying when looking at the entry "fichier" (file).

It is worth pointing out that some collocates have complex meanings and can be classified into more than one class. This is the case with *exporter* (export) that conveys both the meanings of transforming and transferring. We thus placed the verb in two classes (**TRANSFORMER** and **TRANSFÉRER**) thinking that users might access collocates from these different access points.

4.2 Browsing collocations in the dictionary

As mentioned previously, the DiCoInfo is an XML-based dictionary. The classes are naturally declared and organized in a hierarchy that is also modeled as an XML structure suited for such interlocking. At this first stage of the implementation, the hierarchy is limited to four distinct types of classes strictly ordered: a *root*, the generic and intermediate classes (such as **CRÉER/SUPPRIMER**: TO CREATE/TO DELETE; and **Supprimer/Détruire**: To delete / Eliminate), and lastly terminal classes that are the LFs names. Yet, this simple implementation has an essential feature: it allows intermediate and terminal classes to have more than one parent. This characteristic contributes significantly to improve browsing paths as it makes it *a priori* possible to describe parallel access paths to collocations based on different *points of view* (as argued in Section 2); or to classify more accurately collocations that have complex meanings (as exemplified with the *exporter* case in Section 4.1). Figure 1 bellow shows the low complexity of the actual hierarchy and the respective proportion of the different classes (terminal classes are not connected to intermediate ones for clarity).

To ease their management, the inventory and organization of collocation classes that emerge from the grouping analysis are stated as data independently of the dictionary entries (i.e. with the exception of LFs that formalized the links between head words and collocates, no reference to other classes is made within the entries) and the programs that manipulate them. This way, terminologists may easily access and modify at will the organization of the classes to shape the browsing paths faster without having to edit the entries or the programs. The online version of the DiCoInfo (including pages from the search interface) are created by means of XSL transformations of the initial XML dictionary files into HTML pages. While previous versions of the program that generates the pages simply listed the collocations of a dictionary entry in one long table, the present version loads the class hierarchy as an additional data structure along with the dictionary files, and then displays entries within an

outline view (or *tree view*) section that holds the collocations according to the hierarchy. This new section is first presented as an ordinary hyperlink. By clicking on this hyperlink, users open the hierarchy and may select different *branches* (or *nodes*) according to the class names presented and their search needs (as shown in Figure 2). Ultimately, browsing paths reach the terminal classes and short tables are presented with the usual information about collocations.

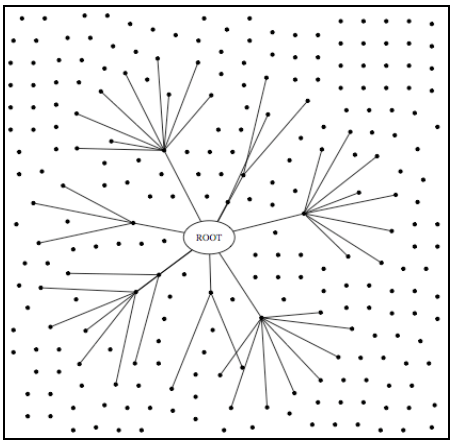

Figure 1: Spring view of the class hierarchy

Figure 2: Outline view of the class hierarchy

5 Usability test

5.1 Background and design

One of the main requirements for the application developed in the framework of this research project is to adapt data access and presentation in order to cater more efficiently for the specific needs of intended users. Hence, we decided to conduct a small-scale pilot-study of the usability of the new data presentation. As explained in more detail in section 3.2, the new presentation is aimed at providing specific assistance in text production situations in L1 or in L2, in which users know the meaning of a collocation but do not recall its components (often verbs with a specialized meaning). Access is then provided to help users find potential candidates for the relevant collocates. This section will attempt to answer the last question asked at the beginning of this paper: Can we assess the usability of the new presentation? By asking two more specific questions: Does it provide a fast and easy access? Can we assess its overall efficiency?

In order to answer these questions, we decided to obtain feedback from the users of the DiCoInfo by means of a test in a controlled experimental environment. The participants were selected from among the intended core users of the DiCoInfo, namely BA-students at Aarhus University, taking the course *Translating and editing texts for corporate websites*.[3] Seven students (out of the entire group of 10) participated in our test. To test the possible impact of the user-profile on the dictionary function (text production tasks in L2 or in L1 respectively), we also decided to gather feedback from 3 French Canadian translation students at the University of Montreal.

[3] The course is designed for Danish students of Business French in the second year of their study programme, and is aimed at enabling them to translate and edit Danish texts from corporate websites into French in the most appropriate way for the French market. As the students need multilingual terminology resources to complete their assignments, they are introduced to resources in the field, including the DiCoInfo.

Prior to the test, the participants were introduced to the functionalities of the DiCoInfo, including a presentation of its theoretical background and resources. The use of the DiCoInfo was illustrated by means of search sessions. The test was designed as a controlled questionnaire with specific instructions for the informants on how to perform the search task and complete the survey after each task (search results and search time), and fell into two distinctive parts: 1. Recording of information retrieval, and 2. Usability assessment. The first part consisted in performing four distinct search sessions mainly aimed at retrieving information from the new presentation in semantic classes. In the case of search session 1 – around the entry *fenêtre* – users were asked to perform search tasks and answer specific questions. The questions were designed to elicit data concerning the accessibility of the forms and the contents of the different information categories addressing *fenêtre*, particularly the new presentation of collocations and their meaning. The participants were asked to record the time spent on retrieving the information, and to assess the usability (access and user-friendliness) of the DiCoInfo. The second part is the evaluation part. It was designed to generate and elicit both quantitative and qualitative data (questions and answers, comments, explanations and suggestions concerning the information found in the DiCoInfo).

5.2 Results: Usability assessment and access and retrieval time scores

Immediately after each of the four search sessions the informants completed a survey in which they evaluated both the information they had found in the DiCoInfo and the ease of access to this information. The evaluation consisted in providing answers to several questions, of which the following two are of immediate concern to us: Was the information easy to understand? Was the information easy to find?

The numbers in each row in Table 4 below show the distribution of evaluation for each of the four search sessions on a 1-4 scale, 1 representing the highest degree of satisfaction and 4 the lowest. As each of the 10 informants answer two questions concerning every session, the total number of answers for each session is 20.

	SATISF. CAT 1	SATISF. CAT 2	SATISF. CAT 3	SATISF. CAT 4
SEARCH SESSION 1	3	14	3	0
SEARCH SESSION 2	0	19	1	0
SEARCH SESSION 3	0	11	9	0
SEARCH SESSION 4	9	11	0	0
% (rounded)	16%	69%	15%	0%

Table 4: Usability assessment

The majority of participants (69%) evaluate the usability of the DiCoInfo as satisfactory; none of them express strong dissatisfaction, while two minority groups express strong satisfaction (16%) or moderate dissatisfaction (15%). The results point to a high degree of overall perceived satisfaction (83%), but also indicate that there is room for improvement of overall usability. Table 5 below shows the highest, lowest, and average time scores for each of the four search sequences:

	HIGHEST TIME SCORE	LOWEST TIME SCORE	AVERAGE TIME SCORE
SEARCH SESSION 1	7 min.	0.5 min.	2.8 min.
SEARCH SESSION 2	7 min.	0.5 min.	2.5 min.
SEARCH SESSION 3	5 min.	1 min.	2.3 min.
SEARCH SESSION 4	4 min.	0.5 min.	2.8 min.

Table 5: Access and retrieval time scores

The time scores are flatly distributed across the search sequences in each of the three categories. There is, however, a striking gap between the highest and the lowest scores, as a few users (the ones expressing low satisfaction) spent about 14 times as much time as the fastest users.

5.3 Qualitative comments

The results of the analysis of the qualitative comments can be summarized as follows:

- Positive: All users emphasize the wealth and high quality of information, and acknowledge the usability of the DiCoInfo as a professional reference work for text production assistance.
- Negative: Almost all users stress the fact that they find it difficult to navigate within the articles, and express their experience in typical statements like "*c'est un peu difficile de s'y retrouver*" (= it's a bit difficult to find one's way), "*on s'y perd facilement*" (= you can easily get lost). Users obviously take the wrong path and need to backtrack. A case in point is the presence of two or more distinctive term records (*interface*), which adds to the difficulty of navigation.
- Suggestions for improvement: The use of distinctive colors to highlight sections, a clearer graphical layout, faster reloading of pages, and an even more efficient search-engine are some of the suggested features.

5.4 A paradoxical situation

Despite being a modest pilot-study, the test and feedback from users indicate that users with a semi-expert profile do appreciate the new resources. Access and retrieval is successful for most of them, provided they have been instructed in the use of the resource. There are, however, a few reservations to the positive output: some users fail to retrieve the information or spend much more time doing it; most users express frustration with the insufficient user-friendliness of navigation and data presentation. In fact, time scores in absolute figures (average time being almost 2.5 minutes!) reveal that speed of search and retrieval is influenced by presentation constraints and problems in understanding the rationale behind it. In short, the test reveals a paradoxical situation. On the one hand, users acknowledge the value of the information and perceive the DiCoInfo as a powerful instrument. On the other hand, using the DiCoInfo appears to be time consuming and demanding. This brings up the question of the lexicographical return on investment. We believe this return to be high, but wish it was even higher. There is still work to be done to achieve better output from the new features. One way to do this could be the integration of better, interactive instructions.

6 Conclusion: Status, future work and challenges

Up till now, approximately 300 LFs have been classified into wide-ranging semantic classes (generic and intermediate). Most represent generalizations over recurrent relationships in the DiCoInfo. Still, more LFs (most of them being non-standard LFs) need to be sorted out and organized within the hierarchy. Future work encompasses the selection of appropriate names for classes (i.e. names that clearly indicate what is comprised under it). On the software side, future work includes testing inheritance and triggering mechanisms in the hierarchy, and adapting the search engine to the new search options. In order to improve usability, we plan to add interactive instructions as well as a graphical search interface. Also, in the future it would be interesting to design and conduct usability tests of the DiCoInfo in real text production situations, and include log records of the number of clicks. This would generate an even more accurate picture of the overall usability of the DiCoInfo and help us in the ongoing development of user adaptive data presentation and access solutions. Finally, the work will have to be adapted to English and Spanish versions of the dictionary.

References

Dictionaries

Benson, M., I. Benson, I. & R. Ilson. 1997. *The BBI dictionary of English word combinations*. Amsterdam/Philadelphia: John Benjamins.

Binon, J., S. Verlinde, J. Van Dyck & A. Bertels 2009. *Dictionnaire d'apprentissage du français des affaires. Dictionnaire de compréhension et de production de la langue des affaires*. http://www.kuleuven.be/ilt/blf [last accessed: 16 June 2009]

DiCE = Alonso Ramos, M. 2011. *Diccionario de Colocaciones del Español*. Universidade da Coruña. http://www.dicesp.com [last accessed: 29 June 2011]

Hill, J. & Lewis, M. 1997, 2002. *LTP Dictionary of selected collocations*. Boston: LTP.

L'Homme, M. C. 2011. *Le DiCoInfo. Dictionnaire fondamental de l'informatique et de l'Internet*. http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi [last accessed: 15 March 2011]

Le Fur, D. (dir.) 2007. *Dictionnaire des combinaisons de mots*. Paris: Le Robert

McIntosh, C. (dir). 2009. Oxford Collocations Dictionary for Students of English. 2nd Edition. Oxford: Oxford University Press.

Mel'čuk, I. et al. 1984-1999. *Dictionnaire explicatif et combinatoire du français contemporain*. Montréal: Presses de l'Université de Montréal.

Other references

Alonso Ramos, M. 2004. Elaboración del Diccionario de colocaciones en español y sus aplicaciones. In Bataner, P. & J. de Cesaris (eds.). *De Lexicographia. Actas del I Simposio internacional de Lexicografia*, 149–162, Barcelona, IULA-Edicions Petició.

Charest, S., É. Brunelle, J. Fontaine & B. Pelletier. 2007. Élaboration automatique d'un dictionnaire de cooccurrences grand public. In *Actes de TALN 2007*, 283–292.

Cinkova, S. & P. Hanks Wyndham. 2010. *Validation of Corpus Pattern Analysis – Assigning pattern numbers to random verb samples. Annotation Guidelines*. http://nlp.fi.muni.cz/project/cpa/CPA_valiman.pdf

Jousse, A.-L. 2007. Organisation des fonctions lexicales pour un meilleur accès à l'information dans le DiCo. In Loiseau, M. et al. (eds.). *Autour des langues et du langage: perspective pluridisciplinaire. Papiers sélectionnés du Colloque International des Étudiants Chercheurs en Didactique des Langues et en Linguistique*, P.U.G.

Jousse A.-L. 2010. *Modèle de structuration des relations lexicales fondé sur le formalisme des fonctions lexicales*. PhD dissertation. Département de linguistique et de traduction, Université de Montréal & UFR de linguistique, Université Paris Diderot (Paris 7).

Jousse A.-L., A. Polguère & O. Tremblay. 2008. Du dictionnaire au site lexical pour l'enseignement/apprentissage du vocabulaire. In Grossmann F. and S. Plane (eds.). *Lexique et production verbale. Vers une meilleure intégration des apprentissages lexicaux*, Coll. Éducation et didactiques, 141-157. Villeneuve d'Ascq: Presses Universitaires du Septentrion.

Klosa, A. U. Schnörch & P. Storjohann. 2006. Elexiko — a lexical and lexicological, corpus-based hypertext information system at the Institut für deutsche Sprachemannheim. In *Proceedings of the 12th EURALEX International Congress*. 425-430, Turin, 6-9 September.

L'Homme, M.C. & P. Leroyer. 2009. Combining the semantics of collocations with situation-driven search paths in specialized dictionaries, *Terminology* 15(2), 258-283.

L'Homme, M.C., P. Leroyer & B. Robichaud. 2010. Advanced Encoding for Multilingual Access in a Terminological Database – A Matter of Balance, In *Terminology and Knowledge Engineering Conference 2010. Presenting Terminology and Knowledge Engineering Resources Online: Models and Challenges*, 12-13 August, Dublin.

Mel'čuk, I, A. Clas & A. Polguère. 1995. *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve (Belgique): Duculot / Aupelf - UREF.

Polguère, A. 2000. Towards a Theoretically-motivated General Public Dictionary of Semantic Derivations and Collocations for French. In *Proceedings of the 9th Euralex International Congress,* 517–527, Stuttgart, 8-12 August.

Polguère, A. 2003. Collocations et fonctions lexicales : pour un modèle d'apprentissage. In Grossmann F., and A. Tutin (eds.). *Les Collocations. Analyse et traitement*, 117-133, Amsterdam: De Werelt.

Tutin A. 2010. Les collocations dans les dictionnaires monolingues spécialisés de collocations. *Actes du 2e Congrès Mondial de Linguistique Française* (CMLF-2010).

Verlinde, S., T, Selva & J. Binon. 2006. The Base Lexicale du Français (BLF): A Multifunctional Online Database for Learners of French. In *Proceedings of the 12th Euralex International Congress*, 471–483, Turin, 6-9 September.