The *Base lexicale du français* (BLF):
A Multifunctional Online Database for Learners of French

Serge Verlinde
Thierry Selva
Jean Binon
K.U.Leuven – Groupe de recherche en lexicographie pédagogique (Grelep)
Dekenstraat 6
B-3000 Leuven
Belgium
serge.verlinde@ilt.kuleuven.be

Abstract
Powerful lookup capabilities are included in almost every electronic dictionary. However, improvements are still conceivable in the exploration of the lexical information within the dictionary and in the integration of the dictionary into other applications. In this paper, we illustrate the most salient functions of the *Base lexicale du français* (BLF), an online lexical database for general French.
The major features of this interactive database are the integration of a dictionary (DAFLES), automatically generated exercises (ALFALEX) and corpus applications (CATS) in order to create a powerful learning environment for learners of French. The DAFLES focuses on the learner's needs by providing separate access paths according to receptive or productive needs. ALFALEX trains the language learner on global lexical knowledge as it is described in the DAFLES. The corpus applications enrich the lexical description by providing the (comparative) combinatorial profiles of words as well as examples and sentences for the exercises.

1 Introduction

So far, electronic versions of (French) paper dictionaries have increasingly improved the accessibility of information and the overall readability of the articles. Full text search (Le Petit Robert), direct access to word combinations (Le Grand Robert) and lookup possibilities including vague spelling or pronunciation (Le Petit Robert) are only some of the tools provided by the major commercial French dictionaries on CD-Rom. Furthermore, the TLFi (*Trésor de la langue française informatisé*, Pierrel et al. 2004), as an example of an online version of a paper dictionary, incorporates even more powerful lookup capabilities, including queries which can be launched on various 'objects' in order to find, for instance, examples of a given word used by an author. However, Leech and Nesi (1999) still point out other ways of improving electronic dictionaries: by varying access points, by varying search or access paths and by enabling the interaction between the dictionary and a CALL application. As far as we know, such improvements have only been partially implemented in large scale dictionaries up till now.

The BLF (*Base lexicale du français,*[1] screen 1) is an attempt to create a full scale online learning tool for general French integrating three components: an electronic dictionary turned into a powerful exploration tool (DAFLES, *Dictionnaire d'apprentissage du français langue étrangère ou seconde*), a learning environment directly linked to the dictionary (AL-FALEX) and tools for corpus exploration (CATS, *Corpus Analysis Tools*). For Business French, another electronic dictionary has been made available: the DAFA (*Dictionnaire d'apprentissage du français des affaires*, Binon et al. 2005).

Screen 1. Homepage of the BLF

We have already had the opportunity to present some aspects of the BLF and more particularly of the DAFLES at previous Euralex-conferences (Binon et al. 1998, Selva et al. 2002,

---

[1] The BLF is still under construction. The present version is available at www.kuleuven.be/ilt/blf. Access is free, but some functions are password protected and thus reserved to K.U. Leuven students and staff. For other functions, the output is limited to a certain number of records or articles.

Selva and Verlinde 2002, Verlinde et al. 1998 and 2004). In this paper we want to present the overall architecture of the software and some of its salient characteristics.

2 Architecture of the software

The BLF is a large MySQL database. The user has interactive access to the database by means of PHP scripts, which allow him to run queries against the database and which display the result of the query on the screen (see also the *ELDIT*, the *Dictionnaire des collocations*, the *DiCouèbe* and the *DICE*, which are comparable online dictionaries).

2.1 DAFLES (Dictionnaire d'apprentissage du français langue étrangère ou seconde)

The key component of the BLF is the DAFLES, a learners' dictionary of general French. The DAFLES is a semi-multilingual dictionary with English and Dutch access (other languages, such as German, Italian and Spanish are planned, on the model of the DAFA).

According to whether his needs are receptive or productive, the dictionary user follows different access paths (for screen captures, see Selva et al. 2002).

For instance, starting from a French word, so in receptive mode, the user gets a link to the article explaining the word and, at the same time, to all word combinations containing this specific word. If the user simply wants to discover the meaning of the word, the link leads him to a screen with short definitions, synonyms, domain indications and translations. However, comprehension is often not a matter of isolated words, but of word combinations. The word combinations displayed appear in a specific order: word combinations whose meaning is not compositional and which are thus more difficult to understand precede those whose meaning is compositional. In both categories, more frequent combinations are listed before the less frequent ones.

In productive mode, starting from a word in English or Dutch, the user accesses a screen with long, Cobuild-like definitions, as well as syntactic constructions and examples. In this mode, word combinations are classified by simplified lexical functions, based on Mel'čuk's research (1995). In addition to the word combinations listed, we provide a combinatorial profile (Blumenthal, forthcoming) or a word sketch (Kilgarriff et al. 2004) for a lot of words.

Synonyms, antonyms and hyperonyms/hyponyms are displayed in WordNet-like networks. The definitions render the relations between the words, more or less as in the Longman Language Activator. For some groups of words, the semantic distinctions are very subtle. In order to help the user to catch the exact meaning and the appropriate word combinations of all these words, we have therefore added comparative combinatorial profiles or word sketches for some word pairs (*jour* and *journée, an* and *année* for instance) or even for word groups (*durant, pendant, au cours de* and *lors de* for instance).

Finally, with the "schémas actanciels", the user gets the prototypical words which appear around the verb in subject or object position or in the position of another complement of the verb, thus merging morphology, syntax and semantics (Verlinde et al. 2004).

More specific queries can also be launched against the database, similarly to the TLFi queries. Our purpose being didactic, the queries tend to extract from the database information that is relevant from a learner's point of view (screen 2).
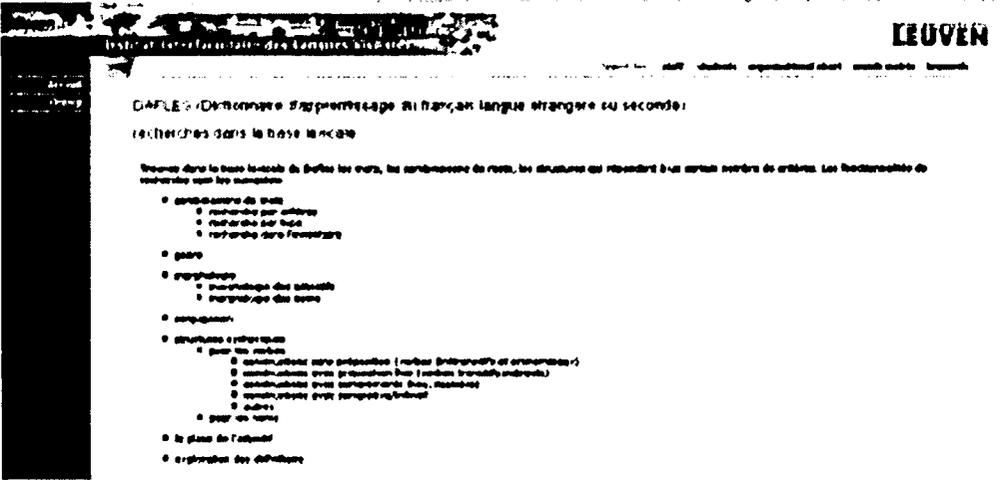
Screen 2. Example of a query

In order, for instance, to retrieve certain specific sets of word combinations, a query combines parameters such as the type of the base of the word combination (noun, verb, ...), the lexical functions, the frequency of the combination, the knowledge domain and the compositionality of meaning. Other queries provided allow the user to explore the gender rules for French (screen 3)
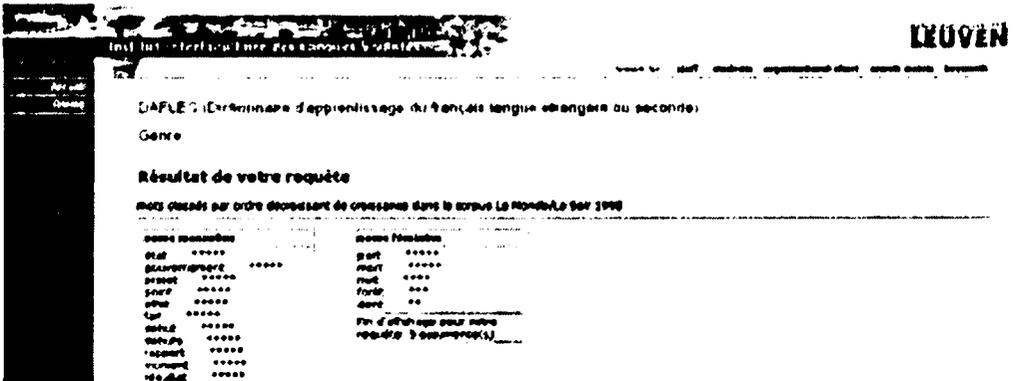
Screen 3. Result of a query: gender of nouns ending in *-t*

or to extract all verbs with the same syntactic pattern: *N1 V N2 à N3, N1 se V à N2, N1 V Vinf*, etc. Among other things, the standardized definitions can also be searched in order for instance to select the words expressing an action, by searching on the string *"est l'action de"*.

The two main differences between the password protected version and the free accessible version of the BLF are the number of records shown and their order. In the free accessible version, this number is always limited and the records are sorted in a random order. In the password protected version, there is no limitation and the data are sorted by descending frequency, with an indication of the frequency class to which they belong.

The lexicographic description of the DAFLES is based on a 75 millions word newspaper corpus (Le Monde (F), Le Soir (B), Le Devoir (Qu), 1998-1999). We also use this corpus to extract examples of word combinations. At this moment, in the password protected version of the DAFLES, the user has direct access to all occurrences of any word combination in the Le Monde corpus. The same application is available in the free accessible version, but examples are taken from a small part of the corpus. The context displayed is limited to a sentence.

2.2 ALFALEX *(learning environment for French vocabulary)*

ALFALEX semi-automatically generates exercises on various items related to words: morphology (verbs, nouns, adjectives), derivation, gender, lexical combination patterns (use of prepositions, word combinations such as support verbs, intensification, fixed expressions, ...), synonyms, "schémas actanciels" and translation to and from French (see www.kuleuven. ac.be/alfalex/Alfalex.pdf for more details). To generate all these exercises, we use the information contained in the DAFLES database: word families indications are used for to practise derivations, syntactic description added to the examples in the DAFLES for the exercise on prepositions, classes of nouns and adjectives with the same morphological patterns for the exercise on morphology, etc. All the sentences used in the exercises have been extracted (semi-)automatically from the corpus by simple pattern matching (for the exercise on prepositions for example), or by more complex NLP-processing, e.g. identifying only those variable words whose ending is determined by the head noun, for the exercise on gender agreement (screen 4).

Screen 4. Exercise on gender agreement

Given answers are compared to the information in the dictionary, making balanced correction possible in some cases. For instance, when a user gives a present tense instead of a

future of the same verb, the environment recognizes the form and considers the answer as partly correct.

Feedback is also generated automatically: the user can click on the links added on the correction page to get a summary of information extracted from the DAFLES. For the exercise on support verbs, for example, all combinations with the support verb can easily be accessed on an additional screen.

2.3 Corpus applications

As we mentioned previously, our 75 million words corpus, which is used during the dictionary making process, provides plenty of examples of word combinations in the DAFLES and is the source of the sentences found in ALFALEX.

The BLF also integrates an inhouse software package (CATS, *Corpus Analysis ToolS*) originally designed to aid teachers to compile and to analyse their own corpus. In the password protected version of the BLF, the original and lemmatized versions of the Le Monde corpus are available for consultation through a lot of common corpus analysis tools: search functions on words, lemmas, word categories. As mentioned above, word profiles, based both on the log likelihood measure and the Z-score, can also be generated. Consequently, the dictionary user is able to explore raw linguistic data in addition to those given in the DAFLES.

As for word combinations, CATS produces a graphical interface (screen 5) and an entire list of collocates with their log likelihood or Z-score measure, which may be saved for further use. This application is clearly meant more for research purposes.

Screen 5. Z-score and frequency of combinations with *durant*

The same data are used in another, more pedagogical application helping the learner to encode correctly his message. This output is very similar to the word sketch output (Kilgarriff et al. 2004): the collocates of a word are listed by grammatical category and sorted by Z-score (screen 6).

Screen 6. Collocates of the word *condition*

By clicking on the link, the user opens the article of the word in the DAFLES or gets a simple translation (> *trad.*) provided by the free InterGlot dictionary.

2.4 The BLF as decoding tool

The BLF may also be used as a decoding tool. The present version analyses sequences of up to ten words, providing a morphological analysis (word forms > lemmas) and detecting possible word combinations. Grammatical words are not considered (screen 7).

Screen 7. The BLF as decoding tool

If a lemma appears in the DAFLES, there is a link to the full article. For the other words, there is a link to the InterGlot translation dictionary, translating the word into the language chosen by the user (Dutch, English, German, Spanish or Swedish) and to the French version of Wikipedia. All detected word combinations are also clikable.

3 Use of the software

We do not have precise information on the use of the DAFLES. Nevertheless, we see that the use of our other online dictionary (DAFA: Business French), which is very similar to the DAFLES, increases by some 20% every year. As the other applications of the BLF are too recent we do not yet have reliable statistics on their use.

Over 2000 students use ALFALEX at our university. Every year, more than 1.000.000 exercises are made. All the students' answers are saved in our database and are, in turn, used to

feed a help function for the more difficult exercises (collocations, synonyms, "schémas actanciels"): the most frequent answers appear as multiple choice answers for each item (screen 8).
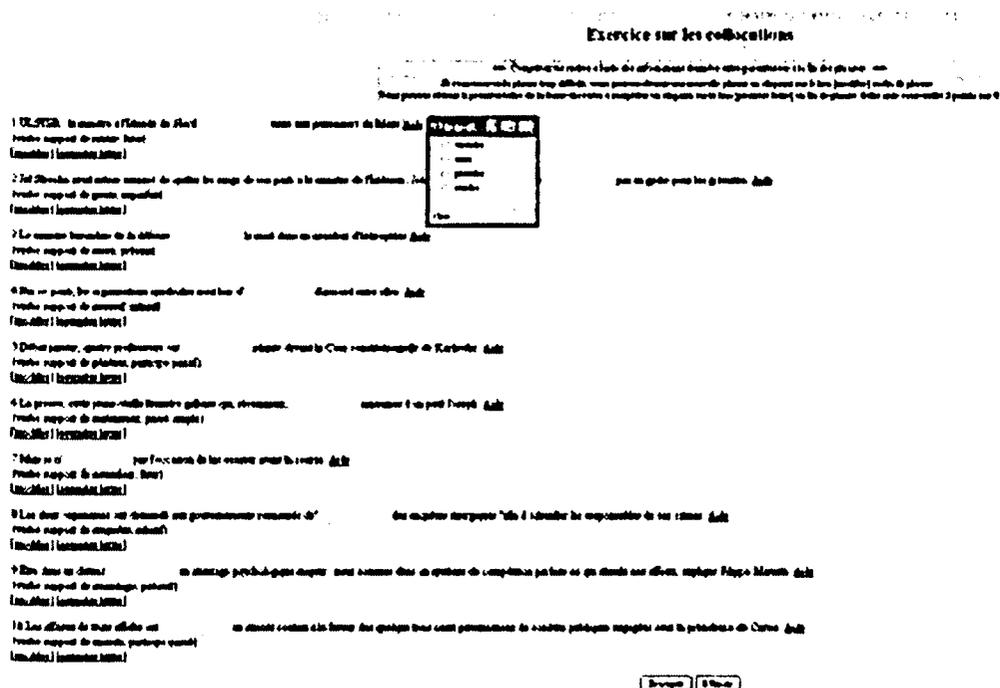
Screen 8. Exercise on collocations

All these answers also help us to improve our course books, since they provide us with reliable feedback about the most common errors of our learners.

CATS is currently used by language teachers at our institute for projects on Legal and Medical French, for Business English and for Dutch.

4 Conclusions

As we have seen the BLF offers a wide range of powerful tools which are useful for lexicographers, teachers and researchers as well. A multifunctional online database for learners of French is all the more welcome, and even necessary because for the French language there do presently not exist any equivalent tools as the ones which are available to teach and study English.

We hope therefore our BLF will contribute to an improvement of both learning and teaching French as a foreign and/or a second language.

References
A. Dictionaries
Cobuild. Sinclair, J. M. (ed) (2003), *Collins Cobuild Advanced Learner's English Dictionary*. Glasgow, HarperCollins.
Le Grand Robert. Rey, A. (ed.) (2005), *Le Grand Robert de la langue française*. Paris, Le Robert/SE-JER. (CD-Rom version)
Le Petit Robert. Rey, A. and Rey-Debove, J. (eds) (2001), *Le Petit Robert. Dictionnaire de la langue française*. Paris, Dictionnaires Le Robert/VUEF. (CD-Rom version)
Longman Language Activator. Summers, D. (ed.) (1993), *Longman Language Activator. The World's First Production Dictionary*. Harlow, Longman.

Online dictionaries (links checked on 24-March-2006)
BLF: www.kuleuven.be/ilt/blf
DAFA: www.projetdafa.net
DICE: www.dicesp.com
DiCouèbe: olst.ling.umontreal.ca/dicouebe/main.php
Dictionnaire des collocations: www.tonitraduction.net/
ELDIT: dev.eurac.edu:8081/MakeEldit1/Eldit.html
InterGlot: www.interglot.com
Wikipedia: fr.wikipedia.org/wiki/Accueil

B. Other Literature
Binon, J., Dancette, J. and Verlinde, S. (1998), 'Comment améliorer le traitement des synonymes dans un dictionnaire de langue', in Fontenelle, Th. et al. (eds), *Actes Euralex'98 Proceedings. (Communications soumises à Euralex'98 (Huitième congrès international de lexicographie) à Liège, Belgique)*. Vol. I, pp. 77-86.
Binon, J., Verlinde, S., Selva, Th., Bertels, A. and Van Dyck, J. (2005), 'Quelques innovations récentes dans le domaine de la lexicographie pédagogique. La contribution du Dictionnaire d'Apprentissage du Français des Affaires (DAFA) (version papier et électronique)', *Cahiers de lexicologie* 87, pp. 1-36.
Blumenthal, P. (forthcoming), 'De la logique des mots à l'analyse de la synonymie', *Langue française*.
Pierrel, J.-M., Dendien, J. and Bernard, P. (2004), 'Le TLFi ou Trésor de la Langue Française informatisé' in Williams, G. and Vessier, S. (eds), *Proceedings of the Eleventh EURALEX International Congress*, Lorient, Vol II, pp. 165-170.
Kilgarriff, A., Rychly, P., Smrz, P. and Tugwel, D. (2004), 'The Sketch Engine', in Williams, G. and Vessier, S. (eds), *Proceedings of the Eleventh EURALEX International Congress*. Lorient. Vol I, pp. 105-115.
Leech, G. and Nesi, H. (1999), 'Moving towards perfection: The learners' (electronic) dictionary of the future', in Herbst, T. and Popp, K. (eds), *The Perfect Learners' Dictionary (?)*. Tübingen, Niemeyer, pp. 295-306.
Mel'čuk, I., Polguère, A. and Clas, A. (1995), *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve, Duculot.
Selva, Th., Verlinde, S. and Binon, J. (2002), 'Le Dafles, un nouveau dictionnaire électronique pour apprenants du français', in Braasch, A. and Povlsen, C. (eds), *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002*. Copenhagen, CST. Vol I, pp. 199-208.
Selva, Th. and Verlinde, S. (2002), 'L'utilisation d'un dictionnaire électronique: une étude de cas', in Braasch, A. en Povlsen, C. (eds), *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002*. Copenhagen, CST. Vol. II, pp. 773-783.
Verlinde, S., Dancette, J. and Binon, J. (1998) 'Redéfinir la définition' in Fontenelle, Th. et al. (eds), *Actes Euralex'98 Proceedings. (Communications soumises à Euralex'98 (Huitième congrès international de lexicographie) à Liège, Belgique)*. Vol. II, pp. 375-386.

Verlinde, S., Selva, Th., Binon, J. and Petit, G. (2004) 'Les schémas actanciels dans le dictionnaire: point de convergence entre la morphologie et la sémantique lexicale', in Williams, G. and Vessier, S. (eds), *Proceedings of the eleventh EURALEX International Congress.* Lorient. Vol II, pp. 427-437.