

Resource pooling for technical English learning via lexical access

Valérie BELLYNCK

équipe STG, LGP2
461, rue de la Papeterie – BP 65
38402 Saint-Martin d'Hères - France
Valerie.Bellynck@efpg.inpg.fr

Christian BOITET

équipe GETA, laboratoire CLIPS
385, rue de la Bibliothèque - BP 53
38041 Grenoble Cedex 9 - France
Christian.Boitet@imag.fr

John KENWRIGHT

Cellule TICE, bureau 2.12,
701, rue de la piscine, BP 81
38402 Saint-Martin d'Hères - France
John.Kenwright@inpg.fr

Abstract

The first stage of this project aims to facilitate technical English teaching, especially for vocabulary acquisition. We are pursuing two immediate goals: maximizing positive student contributions, even outside of the classroom, and minimizing teacher intervention. The resulting application will support our investigations on what can entice users to contribute collaboratively towards enriching a bilingual technical lexicon in a fertile teaching context.

1 Introduction

The cost of building respectable bilingual or multilingual dictionaries specialized in a certain technical field is very high if one uses professional lexicographers and terminologists. Even if enough money is available, such professionals are quite difficult to find for many domains.

Hence, many projects have been started to create such lexical resources via Internet, by setting up web sites requesting free contributions.

However, it is difficult to entice web surfers to contribute without any kind of reward. This is the specific type of problem that the Papillon project (<http://www.papillon-dictionary.org/>) is encountering (Mangeot-Lerebours 2001, 2003).

One solution is to offer a service such as the Oki Electric web site (<http://www.yakushite.net>) where free access to the Pensée MT system is offered, in exchange for contributions to bilingual dictionaries (organized in a hierarchy corresponding to domains of interest and associated communities) (Murata 2003).

In our case, we would like to “populate” the Papillon database, by letting students in classes of computer science and English in a French engineering school, contribute specialized terms and their translations (plus if possible definitions and references). Our proposal invites students to contribute dictionary items as part of their English course assignment. The idea, then, is not only to exchange contribution for grades, but more so to stimulate mutual aid, increase motivation, favour

self-learning, attach importance to student implication in their education, create a lasting tool which can accompany them through their working life and finally a common interest and pride in their acquisition of a foreign language.

In the first section, we will explain the teaching and learning context in more detail (students, goals, resources). In the second section, we will explain how to merge access and contribution to the lexical database in this context. In the third section, we will describe the current version of our system and associated contribution-based web site, ITOLDU (Industrial Technical On Line Dictionary for Universities) – extranet version in <http://www.pagesperso.laposte.net/kenwright/ITOLDU>.

In the last part, we will present some ideas on how to induce more contributions from users.

2 Teaching and learning context

2.1 Size and types of classes

At the EFPG engineer school, we train each year about 200 students in 10 groups, 3 years of study for each class. We have to manage different initial English levels, some students having learned English as a second foreign language (LV2). Next year, the ITOLDU web site will be used via the EFPG intranet.

The technical specific fields cover pulp and paper science, fiber chemistry, packaging, rheology, digital printing, and colour management.

As a primary use, for preliminary experiments to evaluate usability, the ITOLDU web site will be accessed via an extranet version by a class of 6 “sandwich course” students doing a technical degree.

The experiment has taken place between 15th May and 30th June 2004, and was divided into:

- 2 two-hour lessons,
- three weeks later, one 3 two-hour lessons,
- and finally, three weeks later, 2 two-hour lessons followed by a final exam.

The interest in testing ITOLDU lies in the imposed spacing between the lessons and the opportunity for students with varying levels of English to

contribute to vocabulary acquisition and share findings with their "community".

In the first test, the specific field was not technical but common, and centered on professional communication, because our EFPG students were away doing internships, so that we experimented the application with a small group of students in economics before the real use this year (2004-05).

2.2 About the vocabulary to be learned

- Learning technical English is heavily sought after by French institutions.
- The most important direction is English – French: the tool should help remembering English terms to express accurate technical concepts.
- The students don't know the technical terms.
- There are probably 10000-20000 terms.
- The basic part is to be learned by all students and represents about 10% (1000-2000 items).
- Each student should choose and learn a small fraction of the remaining 90%.

3 How to merge access and contribution

Human manipulation of digital dictionaries helps users firstly to use new ways of accessing words, and secondly to take their actions into account as "unconscious contributions".

The most important factors seem to be the tightness of integration of contribution of the contributing and learning environments, and the simplicity of both web interfaces.

3.1 Access and contribution

To access words via a dictionary, people can start from synonyms they have in their head, look up their definitions, choose the one which seems the nearest, and then move again to words used in that definition. But one can also begin to read the dictionary from any page, trying to find some related idea ("linear" access).

To access words through a discussion with someone else, one can begin by expressing an idea, and then stop if that person can't find the word, ask people around to help find an expression or a word that could take the place of the sought after expression, and continue.

To access a digital dictionary, one is usually limited to entering a lemma (or wordform if there is a "lemmatize" option), and to filtering via a small number of constraints (part of speech/clause, domain, variety such a GB/US). The usual methods are already closed to the book access, but without its "linear" extension, which would anyway be limited by the screen-window. To extend the access to more "human" ways, there are two problems. Firstly, how to express the request (how

to specify the word looked for)? Secondly, how to solve this problem and transcribe the request in the digital access?

A proposal for a few modalities of access has been presented in a paper on "Sensillons for the Papillon project" (Bellynck, 2002).

The other point is to find how to transform the passive use of a digital dictionary to an active contribution to its creation. Ease of use is even more important than for access, but it is not enough. Effort for contribution must be minimized and some kind of reward seems to be necessary.

3.2 Teaching and learning context

The context of English learning allows us to use the same experimental contexts for variants of experiments. Basic vocabulary needs are covered as well as specific technical ones shared by some communities.

The teaching learning context leads us to divide the vocabulary in domains of use (business, basic, or technical English for different specialities), left to the teacher's choice.

Asking students to look for the French translation(s) of an English technical term may reveal the need for a strategy different from that used in the case of basic English, particularly in our case, where French students don't yet know the technical terms in their own language well enough.

The current version could be used with other languages, but our learning context concerns only the direction from French to English.

To investigate the modalities of access, we need voluntary and motivated users. In a learning context, the teacher can simply motivate the user of a digital dictionary to contribute by taking into account the quality of the contribution with specific bonuses awarded in return for the student's evaluation of a given translation. But the teacher often can't spend a lot of time checking up on each contribution of each student: the work is in addition to normal working hours. Our solution will be to let the system take up this function.

3.3 Outline of our project

We present now the first stage of a project that aims to help technical English teaching. We pursue two immediate goals: maximize student positive contributions, even out of courses, and minimize teacher intervention.

The idea is simple: through the English courses and between two courses, each student has to collect or create the lexical data for his own digital dictionary. S/he can choose from existing propositions that s/he finds correct or create her/his own proposition. Selecting an existing proposition generates a vote for the student who has created it.

The resulting application should help us to investigate on what can help users to contribute to collaborative lexical technical thesaurus in the fertile teaching environment.

In the larger project, we want to take advantage of convergent ideas that all entice to favour lexical-user contributions.

4 The ITOLDU system

4.1 Functions of the first version

Recall that, in this first stage, we are pursuing two goals: maximize student positive contributions, even out of courses, and minimize teacher intervention.

The idea is simple: through the English courses and between two courses, each student has to collect or create the lexical data for her/his own digital dictionary based on texts or other sources given by the teacher. The student can also add other words or findings s/he comes across in her/his own pursuit of language acquisition. S/he can choose from existing propositions that s/he finds and correct or create her/his own proposition. Selecting an existing proposition generates a vote for the student who has created it.

4.2 Teacher side of ITOLDU

ITOLDU offers teachers the possibility of supervising student groups, encouraging involvement thanks to bonus marks, and livening up vocabulary via playful word hunts.

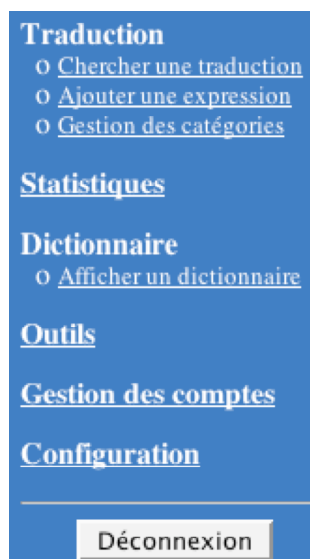


Figure 1: teachers' summary
in Figure 4 ("Statistiques", "Afficher un dictionnaire"). Teachers never have to look inside the source of a html page (or worse in code!).

Figure 1 shows the summary of a teachers' session. One can customize general web service properties (title of the site, language), broadcast learning things to do, contributing to the digital dictionary's construction (search a translation, add a new expression and create new technical domains – called "categories"), manage student groups ("Gestion des comptes"), and look at each student or class-

room contribution shown in Figure 4 ("Statistiques", "Afficher un dictionnaire"). Teachers never have to look inside the source of a html page (or worse in code!).

4.3 Student side of ITOLDU

ITOLDU allows students to gather words or expressions, and to contribute consciously with a proposition of translation or unconsciously with a selection of someone else's translation.

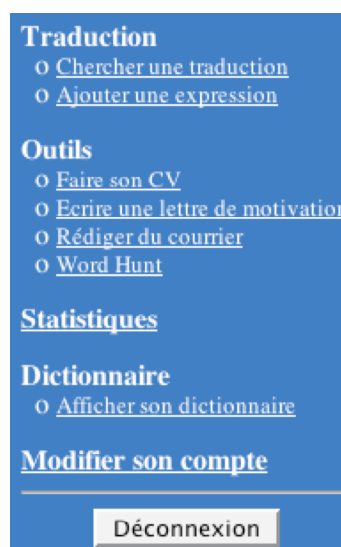


Figure 2: students' summary

When a student connects to her/his own digital dictionary, s/he finds a summary (Figure 2) to access the digital dictionary (search translation and add a new expression), use the teachers' prepared "to-do" tools ("Outils": CV, application letter, word hunt...), look at her/his own statistics, print the current digital dictionary (Figure 5).

4.4 Scenario

Let us imagine that a teacher will prepare her/his course for a classroom and create groups and log-ins. S/he will then give the students some technical English text to study, which includes unknown technical words and expressions. Students will be shown how the ITOLDU tool works, how contributions affect part of their final grade and the concept of sharing knowledge and mutual aid. The teacher can also include an initial "word hunt" (list of targeted vocabulary) to set the ball rolling and encourage users to regularly check the site so as not to be the last to find a word.

Chercher une traduction	
Mot	moonlighting
Traduction	le travail au noir
Contexte	There is a widespread problem of moonlighting in immigrant populations in Italy which has side-effects on the economy.
Source	invented
Catégorie	business english
Vote	75 % (3/4) Charger le mot

Figure 3: basis search access form

When reading a text, a student can be confronted with an unknown word, s/he uses the ITOLDU search tool (Figure 3). In this first version of the

application, the access form is minimal: one can only enter an expression or the first letters of an expression in the first input field. But this form has been designed to be easily replaced or combined with richer ones later.

If there is no entry for the word or expression, the student can enter a translation proposal, with an example of use, the context where s/he has found it, and its bibliographical reference. Each voluntary contribution is cumulated for the statistics and the grades of each student.

If there are one or more entries for the targeted word or expression, the student can select the one which seems to be the best and add it to her/his own dictionary. This action results in an involuntary or unconscious contribution: a vote for the student who suggested this translation (the author). Each vote is cumulated in the statistics of the author (Figure 4). In this case, “jfk” is the name of a student for test (for John-Francis Kenwright).

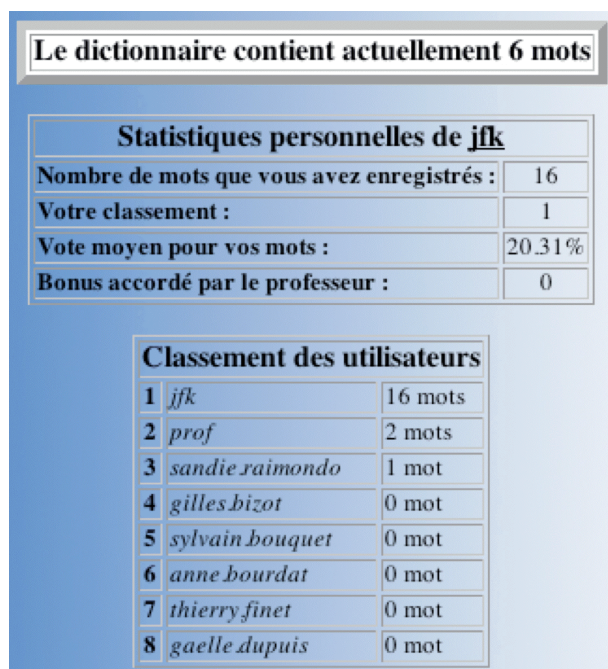


Figure 4: resource pooling statistics

This method of using selections as implicit votes, and further as “unconscious contributions”, is the kernel of the system. As a matter of fact, it will replace teacher mediation. Students can’t enter wrong definitions on purpose, because they would be incorporated in their own dictionaries (Figure 5), and teachers can trace contributions.

For word hunts (shown in Figure 6), the student who finds the word first “wins the game” and has her/his score published on a score board – just like in a computer game.

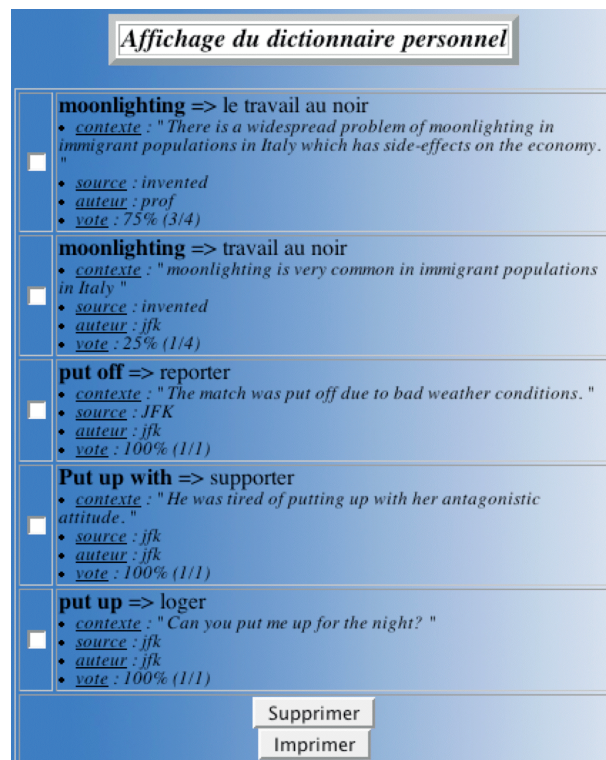


Figure 5: taking over dictionary

Initial experiments were beginning at the same time as this paper was being written, so findings cannot be published here.

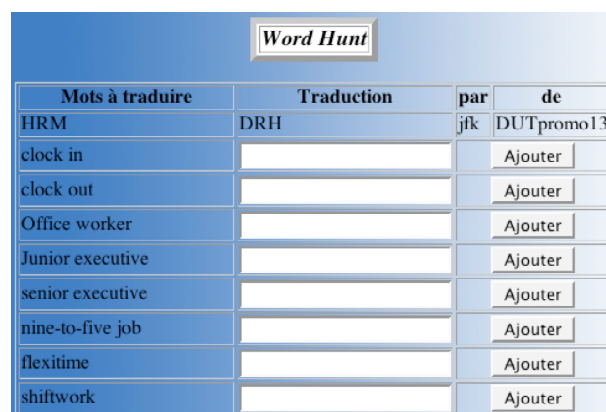


Figure 6: word hunt

5 How to induce more contributions

Other possible ways to induce more contributions are to:

- generalize the “scoreboard” idea so that credits can be shown for each part of the data.
- introduce personalization facilities (automatic or semi-automatic user profiling), so that the system can propose personalized lists of “things-to-do” or new contributions in the user’s domain of interest. For example, the system could remember that a certain user likes to contribute definitions, and propose her/him to complete missing definitions. In Papillon, there are many other types of information to enter, such as pronunciation, example of use, etc., and every user-contributor may have a specific mix of interest in them.

- allow users to self-organize in groups and groups of groups, each group having certain access rights and a profile.
- give users access to tools which can extract potential translation pairs from comparable corpora (texts on the same domain in 2 or more languages, usually not parallel).
- let users contribute directly through an “active reading” interface (where translated words or idioms appear in annotations of text read).
- make the importing environment accessible to users wanting to upload bunches of translation pairs from any format (Excel, Word, FileMaker, XML, etc.). At this moment, to import a dictionary into Papillon, the contributor must put it in XML (with his own tags), and the database manager has to adapt a perl script to convert it into the CXM (Common Dictionary Format) DTD, convert it, and add it to the dictionary collection.
- as the ultimate objective, integrate the lexical contribution function as an add-on (plug-in) in as many applications as possible, used by the general public (text and document processors, spreadsheets, presentation tools, mailers, etc.).

6 Conclusion

We have presented the context use and functions of the ITOLDU system, a web site to help technical English teaching by student resource pooling via lexical access. That context is favourable for gearing users to contribute new terms to the dictionary. In order to ensure quality without asking for teacher’s mediation, we have implemented a voting mechanism.

We will soon report on the first use of ITOLDU with a small class of economics students. The web site will be used for technical English teaching in a French engineer school starting this September. In parallel, we will add more functions to help and entice users, not necessarily students, to contribute lexical data, and conduct experimentations.

7 Acknowledgements

Our thanks go to Cédric Sintès and Sébastien Duvillard-Charvaix for their contribution to the first developments of the project through a student project.

References

- V. Bellynck. 2002. *Bases lexicales multilingues et objets pédagogiques interactifs : Sensillon pour Papillon*. In "Proceedings of Papillon 2002 Seminar", NII, 13 p., Tokyo.
- M. Mangeot-Lerebours. 2001. *Environnements centralisés et distribués pour lexicographes et*

lexicologues en contexte multilingue. PhD in Computer Science, Université Joseph Fourier Grenoble I, 280 p., Grenoble, France.

M. Mangeot-Lerebours. G. Sérasset. M. Lafourcade. 2003. *Construction collaborative d'une base lexicale multilingue, le projet Papillon*, TAL, 44/2, pp. 151-176.

T. Murata, M. Kitamura, T. Fukui, T. Sukehiro. 2003 *Implementation of Collaborative Translation Environment 'Yakushite Net'*. Proceedings of the MT Summit VIII. New Orleans.