

ITOLDU, a Web Service to Pool Technical Lexical Terms in a Learning Environment and Contribute to Multilingual Lexical Databases

Valérie Bellynck¹, Christian Boitet², and John Kenwright³

¹ équipe STG, LGP2

461 rue de la Papeterie - BP 65, 38402 Saint-Martin-d'Hères - France

Valerie.Bellynck@efpg.inpg.fr

² équipe GETA, laboratoire CLIPS,

385 rue de la bibliothèque - BP 53, 38041 Grenoble Cedex 9 - France

Christian.Boitet@imag.fr

³ 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 the ITOLDU 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 is designed to support investigations on what can entice users to contribute collaboratively towards enriching a bilingual technical lexicon in a fertile teaching context. The second stage will be to investigate how to use ITOLDU and similar tools to elicit free (but not necessarily voluntary or even conscious) contributions to the research-oriented, linguistically very rich multilingual PAPILLON lexical database.

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, several 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 an 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, trigger 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 at <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 The Teaching and Learning Context

2.1 Size and Types of Classes with ITOLDU V1 Test

At the EFPG engineer school, we train each year about 200 students in 10 groups with 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). This year, the ITOLDU web site is being used via the EFPG intranet with 200-250 students.

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

For preliminary experiments to evaluate usability, the ITOLDU web site (V1 test) was accessed via an extranet version by a class of 6 mature “sandwich course” students doing a technical degree.

The experiment took place between 15th May and 30th June 2004, and was composed of a total of 14 hours contact teaching and a final 2- hour written and oral exam. Lessons were held every 3 weeks and students had between 2 to 3 two-hours lessons per “contact week”. (The class was composed of students who lived as far away as Paris).

It was interesting to test ITOLDU in this context due to the imposed spacing between 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 centred on professional communication as it was a skill that all of the students in the class needed and could cope with given their varied level of command of the language. The results from this test would serve as a point of reflection for any modifications needed in the long-term and prepare teachers for any trouble-shooting before the generalisation of the tool in 2004-2005.

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 remember English terms to express accurate technical concepts.

- The students do not yet know the technical terms in English and have only recently encountered them in French.
- There are probably 10,000-20,000 terms with which the teacher is not necessarily familiar (whether in French or in English).
- The basic part is to be learned by all students and represents about 10% (1000-2000 items). In reality they know how to use between 150-300 specific words/terms in English, associated with their technical field (paper science) by the time they leave in 3rd year.
- 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

In order 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).

In accessing 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.

In accessing 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 be limited by the screen-window anyway. Extending 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 via digital access? A proposal for a few access modalities of access has been presented in a paper on “Sensillons for the Papillon project” (Bellynck, 2002).

Another point is to find how to transform the passive use of a digital dictionary into an active contribution to its creation. Use-friendliness is but one of the key factors.

Generating the will to participate in this community is also primordial, along with minimising the time required to add contributions. The notion of “reward” seemed necessary from the start. Thus, students were informed that not only would their participation count towards their final mark but also that the quality of their translations and their implication in “voting” for their peers’ contributions would generate a bonus/minus mark at the end of the term.

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 different communities. The teaching-learning context leads us to divide the vocabulary into domains of use (business, basic, or technical English for different specialities). The teacher has the option of adding or deleting categories according to the needs of his/her class(es).

Asking students to look for the French translation(s) of an English technical term may reveal the need for a strategy which is 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 translations from French to English.

In order to investigate the modalities of access, we need voluntary and motivated users. In a learning context, the teacher can simply motivate students to contribute precise translations with specific bonuses. But in reality, due to time constraints, the teacher often can't spend a lot of time checking up on each contribution of every student: the work would be in addition to normal working hours. Our solution from the outset was to let the community take up this function in the full knowledge that the teacher would check a certain number of contributions per students through the year and that wrong translations or rushed voting would lead to low marks and minus scores.

4 The ITOLDU System

4.1 Functions of the First Version

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. Contributing a translation or 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.

4.2 Teacher Side of ITOLDU

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.

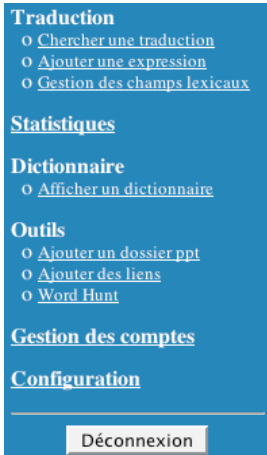


Fig. 1. Teachers' Summary

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.

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, and this measure their own implication or knowledge against fellow classmates, or print the current digital dictionary (Figure 5).



Fig. 2. Students' summary

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 shows the summary of a teachers' session. He can customize general properties (title of the site, language), broadcast learning activities, contribute to the digital dictionary's construction (search for 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"). A heaven-send is that teachers never have to look inside the source of a html page (or worse in code!).

4.4 Scenario

Let us imagine that a teacher prepares her/his course for a classroom and create groups and logins. S/he then gives the students some technical English text to study, which includes unknown technical words and expressions.

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	100 % (4/4) <input type="button" value="Charger le mot"/>

Fig. 3. Basic search access form

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 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). Here, “jfk” is the name of a student, used for testing.

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.

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

Le dictionnaire contient actuellement 241 mots	
Statistiques personnelles de jfk	
Nombre de mots que vous avez enregistrés :	18
Votre classement :	5
Vote moyen pour vos mots :	27,78%
Bonus accordé par le professeur :	0
Classement des utilisateurs	
1 REGA	76 mots
2 BUTY	58 mots
3 thierry finet	32 mots
4 gaelle dupuis	26 mots
5 jfk	18 mots
6 anne bourdat	11 mots
7 hamburger	10 mots
8 sylvain bouquet	7 mots
9 prof	6 mots
10 OKRASA	6 mots

Fig. 4. Resource pooling statistics

The method of using selections as implicit votes, and even more so as “unconscious contributions”, is the kernel of the system.

<ul style="list-style-type: none"> • (un)likely => (peu)probable • <u>contexte</u> : " For example, if you lost something (jewel) during swimming in the sea, you will be unlikely you find it again. " • <u>source</u> : English lesson • <u>auteur</u> : thierry.finet • <u>vote</u> : 100% (1/1)
<ul style="list-style-type: none"> • A going away gift => Cadeau de départ • <u>contexte</u> : " When someone leave a company " • <u>source</u> : English lesson • <u>auteur</u> : thierry.finet • <u>vote</u> : 100% (1/1)
<ul style="list-style-type: none"> • Advanced technician => Technicien supérieur • <u>contexte</u> : " Professional level " • <u>source</u> : Homework • <u>auteur</u> : thierry.finet • <u>vote</u> : 100% (1/1)
<ul style="list-style-type: none"> • assessment => évaluation • <u>contexte</u> : " At the end of the year, we will have an assessment of our english level. " • <u>source</u> : Class. • <u>auteur</u> : gaelle.dupuis • <u>vote</u> : 100% (2/2)
<ul style="list-style-type: none"> • avalanche probe => sonde • <u>contexte</u> : " Use avalanche probe to find people under snow " • <u>source</u> : Avalanche safety • <u>auteur</u> : sylvain.bouquet • <u>vote</u> : 100% (1/1)
<ul style="list-style-type: none"> • Avalanche transceiver => (ARVA) Appareil de Recherche de Victime en Avalanc • <u>contexte</u> : " find people under avalanche with avalanche transceiver " • <u>source</u> : Avalanche safety • <u>auteur</u> : sylvain.bouquet • <u>vote</u> : 100% (1/1)
<ul style="list-style-type: none"> • bank => banque

Fig. 5. Taking over dictionary

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.

Initial experiments were beginning at the same time as this paper was being written, so comprehensive findings cannot be published at this stage. However, in the first test this proved to be a healthy instigator of competitiveness between competing classmates.

perks	Avantages	gaelle.dupuis	DUT promo 13
jobless	Au chômage	gaelle.dupuis	DUT promo 13
Employment agency	Agence de placements	thierry.finet	DUT promo 13
nine-to-five job			Ajouter
Hire and fire			Ajouter
Corporate culture			Ajouter
Long-hours culture			Ajouter
Casual Friday			Ajouter
going rate			Ajouter
cash in hand			Ajouter
job with scope			Ajouter

Fig. 6. Word hunt

4 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, examples 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 two 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.).

5 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 have first used ITOLDU with a small class of economics students. The web site is used for technical English teaching in a French engineer school during this academic year with 200-250 students. In parallel, we will add more functions to help and entice users, not necessarily students, to contribute lexical data, and conduct experimentations.

With this kind of tool, we are implementing one of the two possible solutions to the nagging problem of enticing users to contribute. As it is in practice not possible to get voluntary and free contributions, one can try to get either voluntary and rewarded contributions, or free and involuntary (or even forced!) contributions. The first solution is that implemented in yakushite.net, where users contribute because they use a

free translation support tool where their lexical contributions become quickly active. The second solution is implemented in ITOLDU and can be generalized to any situation where students are learning information of interest, and where teachers are looking for a tool to alleviate their work.

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, and to Mathieu Mangeot and Gilles Sérasset for designing and implementing PAPILLON.

References

1. V. Bellynck (2002). *Bases lexicales multilingues et objets pédagogiques interactifs : Sensillon pour Papillon*. In "Proceedings of Papillon 2002 Seminar", NII, 13 p., Tokyo, July 2002.
2. 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.
3. M. Mangeot-Lerebours. G. Sérasset. M. Lafourcade (2003). *Construction collaborative d'une base lexicale multilingue, le projet Papillon*. TAL, 44/2, pp. 151-176.
4. T. Murata, M. Kitamura, T. Fukui, T. Sukehiro (2003). *Implementation of Collaborative Translation Environment 'Yakushite Net'*. Proceedings of MT Summit VIII. New Orleans, Sept. 2003.
5. Naoyuki Tokuda and Liang Chen (2001). *An Online Tutoring System for Language Translation*. IEEE Transactions on Multimedia, vol. 8, no. 3, pp. 46-55, July-September 2001.