

Bilingual Lexical Data Contributed by Language Teachers via a Web Service: Quality vs. Quantity

Valérie Bellynck, Christian Boitet, and John Kenwright

Abstract—IToldU is a light web service which, in its first year of use for teaching technical English in French engineering schools, has enabled the contribution of just over 17000 English terms in about twenty technical domains. These terms are associated with their French translations (95% of which are correct) and examples of use (about 85% correct). In the second year, emphasis has been on quality rather than on quantity: about 6000 high-quality entries have been contributed by the same number of students and classes. Some desirable extensions are in progress, e.g. to add English when this language is not included in the original language pair, and to synchronize with off-line contributions prepared on a PDA or a hand-held calculator.

Index Terms—Collaborative dictionary construction, examples of use, technical English teaching.

I. INTRODUCTION

THE collaborative construction of free lexical resources has been hampered by the difficulty of obtaining many individual small and voluntary contributions. IToldU (Interactive Technical On-Line Dictionary for Universities) is a light web service which can be used for the collaborative construction of a bilingual lexicon by a small community (typically, a group of students) while learning a foreign language in technical or specific domains. Contributions are freely offered, but are also constrained in that part of the students' English grades are computed by IToldU itself.

For the first two authors, the initial objective in building this site was to collect the produced lexica in order to populate the multi-usage multilingual lexical database (MLDB) Papillon (see <http://www.papillon-dictionary.org/>). For the third author, an English teacher of ICTE (Information and Communication Techniques for Education) at INPG (Institut Polytechnique de Grenoble), the objective was to improve the teaching of technical English vocabulary to French engineering students.

Manuscript received November 25, 2008. Manuscript accepted for publication August 15, 2009.

Valérie Bellynck is with équipe STG, LGP2 461 rue de la Papeterie, BP 65, 38402 Saint-Martin-d'Hères, France (e-mail: Valerie.Bellynck@efpg.inpg.fr).

Christian Boitet is with équipe GETA, laboratoire CLIPS, 385 rue de la Bibliothèque, BP 53, 38041 Grenoble Cedex 9, France (e-mail: Christian.Boitet@imag.fr).

John Kenwright is with Cellule TICE, bureau 2.12, 701 rue de la Piscine, BP 81, 38402 Saint-Martin-d'Hères, France (e-mail: John.Kenwright@inpg.fr).

In its current state, IToldU addresses mainly the instructional objective rather than the lexicographical one. Moreover, its use has led to a third interesting possibility, that of teaching the structure of simple sentences of English through examples in use: it turns out that students are not satisfied with copying and pasting sentences containing the terms they translate, but prefer to create their own examples.

In the following sections, we will: present IToldU; evaluate its first two full years of use (describing its pedagogical impact on students and teachers and the quantitative and qualitative lexicographical results obtained when varying the desired quality level); and describe plans for increasing contributions, for extending collection to other languages and types of information, and for synchronization with the Papillon online multilingual lexical database.

II. THE ITOLDU WEB SERVICE

A. Teaching Context and Goals

The teaching context is as follows:

- Acquiring and using technical English.
- The most important translation direction is English-French.
- Students don't 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 (either in French or in English).
- The teaching goals of the English courses, over the three years spent in the schools by students, are twofold:
 - The base technical vocabulary that is to be learned by all students represents about 10% (1000-2000 items) of the terms.
 - Each student should choose and learn a small fraction of the remaining 90%.

Students know how to use between 150-300 specific English words or terms associated with their technical field (paper industry) by the time they leave in the third year. Of course, they know many more general terms, and terms in all other domains encountered during their courses (including other technical fields, work placements, themes and skills seen in traditional English classes, job hunting, etc.).

Students receive a point for uploading an entry onto their dictionary (effectively “voting” for it). However, if the entry is wrong, the student will lose a point later. In both cases, IToldU motivates students via the possibility of gaining or losing points. This incentive instills in them a positive learning attitude. Moreover, the publication of the “top ten” best scores on the web site motivates them to participate more and more often, creating a healthy competitiveness among individuals and groups.

E. Teachers

IToldU offers teachers the possibility of supervising student groups, encouraging involvement through the use of bonus marks, and livening up vocabulary acquisition via playful “word hunts“. Fig. 3 shows the summary of a teacher’s session.

S/he can customize general properties (e.g. the title of the site, or its language), broadcast learning activities, contribute to the digital dictionary’s construction (by searching for a translation, adding a new expression and creating new technical domains – called “categories”), manage student groups (“*Gestion des comptes*” – account management), and look at the contribution of each student or classroom, as shown in Fig. 4 and Fig. 5 (“*Statistiques*”, “*Afficher un dictionnaire*” – display a dictionary).

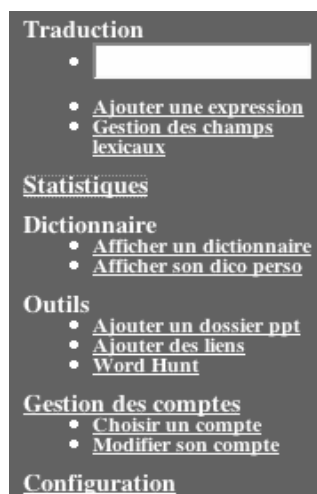


Fig. 3. Teachers' summary.

A particular blessing is that teachers never have to look inside the source code of an HTML page or (even worse!) other program code. Another important point is that the time constraints of the teachers are taken into account: teachers have almost no time to follow students’ work outside the classroom (perhaps 1-2 minutes per student). The use of IToldU should not increase their work time, but if possible reduce it.

That seems to be the case now, as the grading system has been designed to optimize the teacher’s time. During the first few weeks of use by a new group, the teacher systematically goes online and deletes any incorrect words. This supervision encourages rigor at the start of the program.

statistiques du compte courant

Statistiques sur le compte de **REDON**
(poids du dico dans le dico commun : 0.014
Cette année Depuis le début

Statistiques personnelles de REDON	
Nombre de mots que vous avez enregistrés :	90
Votre classement :	122
Nombre de mots que vous avez produits :	60
Nombre de mots importés depuis les mots que vous avez produits :	6
Vote moyen pour vos mots :	1.1%
Nombre de participations à la chasse aux mots :	0
Bonus accordé par le professeur :	0

Classement des utilisateurs de la promotion 2A_06-07 par nb entrées		
1	DOTAL	171
2	HAJJI	84
3	EYBRALY	76

Fig. 4. Resource pooling statistics.

During the second year, evaluations of contributions are scheduled (every five months) in which teachers check a few dictionary samples from each student in their class. Students don’t know which sample will be checked, and are hence motivated to check and improve their entire dictionary. Owing to lack of time and for pedagogical reasons, teachers do not correct mistakes, but simply mark that a translation or an example is wrong. IToldU supports such error marking on fields. Then students must make the corrections before a certain time elapses, or IToldU will subtract the corresponding points.

Fig. 6 shows an example of a “word hunt” screen. “Word hunt” is a challenging but enjoyable part of IToldU for both teachers and students. The first student to find a translation wins a point! Thus students log on as often as possible to see if there are words up for grabs!

III. EVALUATION

A. Pedagogical Aspects

Reactions of teachers and students. The current complete version of IToldU (<http://opus.grenet.fr/itoldu/ITOLDU>) was used for the first time in 2004-05 by all the students of EFPG, an engineering school that is attached to INPG, with a clear positive pedagogical impact. A total of 250 students were involved in the beta test, spread out over the three years of engineering school and one year of professional BA (licence) work. As far as English teaching was concerned, there were 17 groups, 6 teachers, and 1 coordinating teacher (the third author).

IToldU already addresses quite well the need felt by the coordinating teacher for a computer tool improving management of training, teachers’ work, and students’ learning of specialized English technical vocabulary.

B. Dictionary Evaluation (First Year)

1) Quantitative aspect

In the first semester, about 12,000 English-French entries were entered into IToldU by the students, along with about 8,000 usage contexts.

At the end of the academic year, IToldU contained 17,062 English-French entries, and about as many usage contexts (only 157 entries lacked contexts).

2) Qualitative aspect

The second author quickly revised all the contributions of the first year, and about 10% in detail, thereby correcting them. Apart from errors arising from problems in inputting diacritics on the Web, the French translations of English terms are almost all correct. By contrast, 15% to 20% of usage contexts are not examples of use. Following are some details on these two types of contribution.

Translations. 95% of the translations seem correct to us. An interesting point is that only about 30% of the English terms chosen by the students concern a purely technical lexical field, one linked with students' studies (of manufacturing paper pulp, paper, cardboard, color processing, inks, rheology, etc.) while 70% concern "paratechnical" fields, such as business or job hunting, or general English.

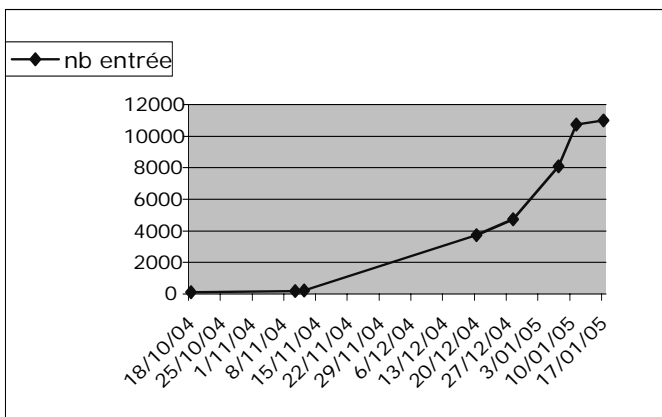


Fig. 7. Evolution of the number of entries in the first semester.

From usage contexts to examples of use. "Contexts" merit some comments. In the mind of the teachers, contexts should be citations of sentences in which the English terms had been encountered. But several unexpected things happened.

Certain students understood that they were being asked for the "domain" of the citation, selected from a list provided by IToldU. One finds for example:

5024	opportunity	possibilité, débouché	society
5025	to put up	ériger, construire	society
5026	to fulfill	accomplir, réaliser	society
5027	fulfilling	profondément, satisfaisant	society
15009	gas-fired	chauffé au gaz	used in paper mill

Others thought that they were being asked for definitions.

15049	a wind mill	une éolienne	an energy-producing facility
15065	a light bulb	une ampoule électrique	energy-related equipment
4632	TCF (totally chlorine free)	sans chlore	stade de blanchiment

The coordinator accordingly modified his description: he asked for "examples in use", and created some himself, putting "invented" in the source field. The students then understood that they, too, could invent examples, and did so. At the level of content, several cases arose:

- Some students created or adapted sentences containing the English terms in question, but in such a way that the word meaning could not be discriminated.

16070	collude	s'associer	they colluded last year
16990	telematics	télématique	it s telematics
16998	darts	fléchettes	he throws the darts
17003	potoling	spéléologie	the potoling is dangerous
17006	chiari-oscuro	clair-obscur	the is a chiaroscuro effect
17026	heir	héritier	you heir to your mother

- At the other extreme, other students used long sentences as examples.

12956	Falsification	Falsification	Some various documents to be protected from counterfeiting and falsification like service vouchers, security label and certificates of authenticity have special features.
12957	service vouchers	Tickets de prestation	
12958	security label	Etiquettes sécurisées	
12959	certificates of authenticity	Certificats d'authenticité	
12960	anti-counterfeiting features	Eléments anti-contrefaçon	
12961	anti-falsification feature	Eléments anti-falsification	

- Many proposals are intended as "honest examples", but are not in correct English.

6619	carriageway	chaussée	the carriageway is destroy by the cars
7073	union	syndicat	an union for help employees
7098	pythonesque	humour absurde	this joke are very pythonesque with his very absurd humor
9183	(to) insulate	isoler	insulating materials can be very useful in electronic

- A small percentage of students vented their frustration by putting "garbage" (silly examples or obscenities) in their examples.

In total, about 15% of the examples are incorrect with respect to content, again not counting input errors, and many

ACKNOWLEDGMENTS

We would like to thank our reviewers for many useful comments, and Mark Seligman for a very detailed revision of the paper and improvement of its language. All remaining errors are of course ours!

REFERENCES

- [1] V. Bellynck, "Bases lexicales multilingues et objets pédagogiques interactifs : Sensillon pour Papillon," in *Proceedings of Papillon 2002 Seminar*, NII, Tokyo, July 2002, 13 p.
- [2] V. Bellynck, C. Boitet and J. Kenwright, "Resource pooling for technical English learning via lexical access," in *Proc. Papillon-04 seminar*, UJF, Grenoble, 30 Aug.-2 Sept. 2004, 5 p.
- [3] V. Bellynck, C. Boitet, and J. Kenwright, "ITOLDU, a Web Service to Pool Technical Lexical Terms in a Learning Environment and Contribute to Multilingual Lexical Databases," in *Computational Linguistics and Intelligent Text Processing (Proc. CICLING-2005)*, A. Gelbukh (Ed.), Springer, LNCS 3406, pp. 319-327.
- [4] M. Mangeot-Lerebours, "Environnements centralisés et distribués pour lexicographes et lexicologues en contexte multilingue," PhD in Computer Science, Université Joseph Fourier, Grenoble I, 280 p., Grenoble, France, 2001.
- [5] M. Mangeot-Lerebours, G. Sérasset, and M. Lafourcade, "Construction collaborative d'une base lexicale multilingue, le projet Papillon," *TAL*, 44/2, pp. 151-176.
- [6] T. Murata, M. Kitamura, T. Fukui, and T. Sukehiro, "Implementation of Collaborative Translation Environment 'Yakushite Net'," in *Proceedings of MT Summit VIII*, New Orleans, Sept. 2003.
- [7] N. Tokuda and L. Chen, "An Online Tutoring System for Language Translation," *IEEE Transactions on Multimedia*, vol. 8, no. 3, pp. 46-55, July-September 2001.