QuickAssist: Reading and Learning Vocabulary Independently with the Help of CALL and NLP Technologies

Peter Wood*

Abstract

Independent learning is a buzz word that is often used in connection with computer technologies applied to the area of foreign language instruction. This chapter takes a critical look at some of the stereotypes that exist with regard to computer-assisted language learning (CALL) as a money saver and an easy way to create an "independent" learning environment. It will also look at what currently available tutorial CALL is able to offer and at how to assess users' independence in this environment. The chapter establishes a working definition of learner independence and shows that tutorial CALL is currently only able to help learners become independent to a limited extent. A paradigm shift in language teaching aimed at promoting learner independence necessitates a shift in the design of dedicated CALL software. As an example of how natural language processing (NLP) technologies can be used to promote independent language learning at an advanced stage, the paper briefly presents QuickAssist, an application which enables learners to work with a German text of their choice using a set of NLP tools, and reports on some findings of a user study.

Keywords: intelligent computer-assited language learning (ICALL), natural language processing (NLP), glossing, independent learning, vocabulary acquisition.

*9 Campus Drive, Room 521, University of Saskatchewan, Saskatoon, SK, S7N 5A5, Canada.
e-mail: peter.wood@usask.ca

How to cite this chapter: Wood, P. (2011). QuickAssist: reading and learning vocabulary independently with the help of CALL and NLP technologies. In S. Thouësny & L. Bradley (Eds.), *Second language teaching and learning with technology: views of emergent researchers* (pp. 29-43). Dublin: Research-publishing.net.

1. CALL - the money saver

In this section, I am looking at tutorial CALL applications, their capabilities and shortcomings, and at ways to use natural language processing tools to help advanced language learners.

While many foreign language departments have been seeing severe budget cuts in recent years, they have been struggling to uphold and possibly even improve the quality of their programs. This is at least the result of a survey among the heads of small to medium sized modern language sections in Canada that my department conducted in 2009 (Julien, Makarova, & Wood, in prep.). While costs for both hardware and software might have been prohibitively high a few years back, the infrastructure to administer CALL at universities is now readily available. Can CALL be proposed as a cost-efficient panacea, and can learners manage some or most of the language learning "independently" with the help of technology? Many instructors still have inflated expectations when it comes to CALL (Holland, 1995). I can corroborate this: on various occasions, I was approached and asked whether it would be a good idea to turn some of our offerings into online courses. That way, we ought to be able to increase enrolments and cut down on costs for instructors. Investing into CALL technologies is not wrong, by any means. Sometimes, however, I have a hard time convincing colleagues that investing into CALL cannot go hand in hand with reducing teaching staff or that it will lead to a dramatic decrease in time they will have to spend on preparing lessons or correcting assignments –at least not if the way languages are taught, learnt, and tested does not change dramatically.

Investing in software that is currently available for universities, by and large, involves considerable licensing fees that have to be covered either by the institution itself or by the students. These fees, for the most part, are not one time investments. Either licences expire after a set period, or the software packages have to be upgraded eventually. In addition, there are fees for maintaining and administering a language lab, servers, as well as costs for personnel that take care of online students, correct and grade their work, answer their questions and monitor discussion boards that are part of many modern course packages.

2. Tutorial CALL - the state of the art

In the past, teaching and learning a foreign language implied that instructors and learners had very specific roles. The instructor would decide on the topic of individual classes, would provide input (knowledge) to learners, who in turn were required to reproduce it on request. This is of course oversimplified. Even practitioners of the grammar translation method (Richards & Rodgers, 2001) would eventually expect their students to form utterances independently.

Tutorial CALL programs, on the other hand, cannot do much more than mimic this idealised traditional language teacher. Higgins (1988) uses the term "magister" to refer to CALL applications of this kind and distinguishes them from applications, which he classifies as pedagogues, that afford their users more freedom in making decisions. Levy (1997) uses the terms "tutor" and "tool" to refer to the functionality of programs. Learners can make use of computer applications to complete a specific task, such as looking up an unknown word in an electronic dictionary. Programs of this sort function as tools. The computer as tutor is what most dedicated CALL software, software developed explicitly for language learning, can be classified. It is designed to function like a human instructor. They can provide students with vast amounts of data and exercises to test their knowledge. As long as the set of possible answers to a question remains within manageable bounds, the computer will also be able to provide students with an adequate feedback. It can correct multiple choice and yes/no types of exercises independently. It can even create some of these exercises automatically (Koller, 2007). As long as the learners are at the beginner level, this range of capabilities enables the computer to act almost as a replacement for a human teacher. There are a number of commercial applications that do exactly this.

Using natural language processing tools, computers are also able to deal with more complex tasks, such as analysing short sentences for syntactic (Heift, 1998; Schulze, 2001) and/or semantic correctness (Bailey, 2008) as long as the domain is restricted or the range of possible errors can be anticipated. A few of these so-called ICALL (intelligent computer-assisted language learning)

applications exist, but only one, Compusensei (Nagata, 1992), is a commercial one. For a detailed discussion of what NLP can accomplish, see Heift and Schulze (2007). However, out of the range of NLP applications available, the language learning software industry has so far only displayed an interest in speech analysis: it is now integrated in many language learning applications where it is successfully used for pronunciation drills. Parsers, corpora and other NLP applications that work robustly have so far not been used in commercial language learning software (Jager, Nerbonne, & Van Essen, 1998; Nerbonne, 2003).

Eventually, however, current computer and software technology reach their limits. While modern parsers are able to work well with complex texts that do not contain any errors, they do not perform reliably when it comes to random learner language (random is used here to indicate a situation where the errors are unpredictable and where the context of the learner input is not predefined). Instructors want their students to be able to construct sentences and texts by themselves. My German students, for example, are required to write short texts after their first six weeks of instruction. At this point, it is necessary for a human corrector to look at the student's text and provide adequate feedback. The amount of "human intervention" that is necessary to correct learner output will increase together with the learner's proficiency level. While tutorial CALL programs can still be used to teach advanced concepts, a human instructor will be necessary to provide feedback on learner language that is random in the above sense. Language learning software that promises to provide a comprehensive course for independent study does so under false pretences, or by providing a learning experience that will sooner rather than later become very artificial because language produced by the learner cannot be adequately analysed.

3. Learner independence

Learner independence, also referred to as learner autonomy, is a widely used buzzword. This has been criticised for a number of reasons. Pennycook (1997)

remarks that the notion of independence or autonomy is a western concept. While independence or autonomy have a positive connotation in western society, it might evoke radically different associations in other cultural contexts that place the society above the individual, in which society acts as a sanctuary for the individual. In such a context, a state of autonomy, of being outside society may well be considered undesirable. Schmenk (2006) traces the term autonomy back to ancient Greek philosophy where it was used to describe a political state. This meaning was later adapted by Enlightenment philosophers like Immanuel Kant. From this perspective, as language learners –because of the role they assume in the context of language instruction– are considered to be in need of guidance and support, it would be wrong to call them autonomous.

Many researchers are aware of the inherent problems of both terms and try to use them cautiously. Nevertheless, it is still necessary to define clearly what independence means in a certain context. With respect to the domain of learning, White (2008) locates independence on three different levels:

- Context/Setting: independence can simply mean that learning takes place without a human teacher, but it can also mean that learners have the freedom to make choices, the freedom to select learning opportunities and the freedom to use resources according to their needs.

- Philosophy/Approach: at this level, independence refers to the roles and responsibilities of teachers and learners in the independent learning context. The teacher's role here is to prepare learners to think about their needs. Learners have to develop the ability to look after their own needs.

- Learner Attributes: learners have to develop the attitudes, beliefs, the knowledge and the strategies to take actions that support their learning process.

The stages of autonomy, that Oxford (2008) adapted from Nunan (1997), will be operationalised here to answer the question of the use of CALL to help in the

development of learner independence. Concentrating on the role of the learner and the tasks/goals, her model can be summarised as follows:

- At stage one, the learner's role is that of a recipient of information. At this stage a dedicated e-learning system would ideally decide on the goals and provide tasks for the learner and would also give an explanation for the choices it makes.

- At the next stage, the learner acts as a reviewer and selector among system-given options, selecting tasks specific to their individual learning needs and determining the order in which they complete these tasks.

- At stage three, learners adapt tasks provided by the system to their needs.

- Next, learners assume the role of creators and invent tasks to reach the goals they have set.

- Finally, learners become teachers themselves, undertake independent research and help others to acquire a second language.

Taking a look at tutorial CALL applications that are available today, it is clear that they are all able to function adequately as the e-learning system described in stage one. It is important to note that only some of them will actually give learners a comprehensive explanation for the selection of tasks and the order in which they are administered. Ideally, learners should be able to expect more than a summary of the grammar points and other topics covered in the lesson.

Most modern systems will also provide learners with a range of choices as well as the option to create individual learning programs by enabling them to select learning materials and to decide in which order they are presented. I am not aware of any tutorial CALL software that fulfils the functions outlined in stages three and beyond. It could be argued that most systems enable students to create individual vocabulary databases and that complex

systems place a large amount of learning elements at learners' disposal, but it is clear that individual needs are hard to predict and that even sophisticated systems would be unable to offer everything. For example, an architect learning Italian in order to read Italian publications on architecture might be able to find a lesson on art history or even texts on some architectural monuments, but nothing that would introduce her to the specific terminology used in her field of interest.

The most important problem that becomes apparent when using Oxford's (2008) model as a benchmark for tutorial CALL software is that there is little that currently available software does in terms of preparing learners to progress to further stages of independence. In order to be able to continue to operate with the limited set of functionality discussed above, the system is forced to keep the learner at stages one or two. Once learners advance past these initial stages, the learning system will either have fulfilled its purpose or become part of a larger pool of resources learners use to proceed in their language acquisition. I would argue that claims made by the producer of tutorial CALL software to provide a comprehensive program that leads learners to a stage of proficiency past an advanced beginner level, are actually misleading. The "Transparent Language German" software is only one example of such software that boasts to teach learners everything to "achieve language proficiency" (details can be found in Wood, 2010b).

Returning to White's (2008) classification of independence, there is another area that developers of tutorial CALL software could take into consideration. While it is clear that CALL applications can not –and will not for some time– uphold the illusion of being an adequate replacement for a human instructor, they can help to develop the kind of skills that White (2008) is pointing to. Including exercises that help develop critical thinking, research tasks, lessons on different learning strategies is certainly possible with currently available technologies. Gradually helping beginning learners to become independent from the system while making them aware of its advanced features (such as dictionaries, Grammar references) would not only help learners, but would also ensure that they would continue to use the application, or at least some of its components in the future as a tool box.

4. Using NLP to promote learner independence

True learner independence, in the sense discussed above, necessarily encompasses a shift in current teaching and learning paradigms. The development of skills for independent learning has to be at the core of every curriculum, but more importantly the traditional roles of instructors and students will have to change dramatically. In a learning situation where learners determine their goals independently and decide on the steps they take to achieve them, there is little room for set syllabi and unified exams. In the remainder of this paper, I am concentrating on NLP software as one example of how CALL can be adapted to promote learner independence at an advanced proficiency stage, and how it can be used in independent learning contexts, with regard to the absence of a human instructor.

As pointed out earlier, NLP applications have been used in tutorial CALL applications, and have been proven to be beneficial at the beginner stage. In this setting, NLP is acting "behind the scenes". Learners are not directly exposed to the parser, nor do they query a corpus or look up words in a dictionary. Exposing learners directly to NLP is not a new idea. Data driven learning has been practised for a number of years. However, the degree of exposure varied in individual approaches. Johns (1991), for example, created learning materials directly for his learners. Learners would then analyse the data in order to explore the semantics of a certain word, common collocations, and syntactic particulars. This approach was considered labour intensive by most instructors (cf. Boulton, 2010), many of whom probably had doubts about the benefits of the method in the first place. Other researchers made the corpus and a query interface available to the students directly. Depending on the underlying technology, using the tools effectively involved a steep learning curve at times.

For QuickAssist, a program that I developed as part of my dissertation, I took an approach similar to the one used in Glosser RuG (Dokter, Nerbonne, Schürcks-Grozeva, & Smit, 1998). In an attempt to follow Colpaert's (2004) software design principles, stipulating that the development has to include a needs analysis, the design, implementation and the evaluation of the software, I researched the

availability of CALL software that was suitable for use with advanced learners of German. When it turned out that there was no such software available I started the development that was guided by these principles:

- the program had to be intuitive and easy to use;
- it should be able to provide a wide range of functionality and give users the freedom to chose the learning objects themselves;
- where possible, freely available NLP resources were to be used;
- the program was to be released under the General Public Licence, making it freely available for learners, instructors and developers.

In QuickAssist, users can cut and paste any German text into the application and access a number of NLP tools via a simple and easy to use interface. This was done in order to enable them to work with the application productively without having to learn a query language or have any other sophisticated background in computing. The design and functionality are very similar to a graphical web browser that most learners are sufficiently familiar with. As can be seen from the images below, users are able to click on individual words and are able to look them up in a German-English dictionary using the 'English' button (Figure 1). They can query a corpus for the word and study it in different contexts with the concordancer function by clicking 'Kontext' (Figure 2). This retrieves sentences containing the selected word from the 300,000 sentences German corpus compiled by the Wortschatz project at the University of Leipzig. Moreover, the application provides information on a word's most common neighbours by selecting 'Nachbarn', on its frequencies with the 'Frequenz' function, and on its morphological structure and word paradigms selecting 'Wortbildung' and 'Formen' respectively. It also enables users to look up names of people, places and cultural artefacts in the German version of the Wikipedia (Figure 3). Contrary to Boulton (2010), I do believe that students can be exposed to the underlying technology, provided that they can work with an intuitive enough interface that lets access NLP tools directly.

Figure 1. Dictionary lookup

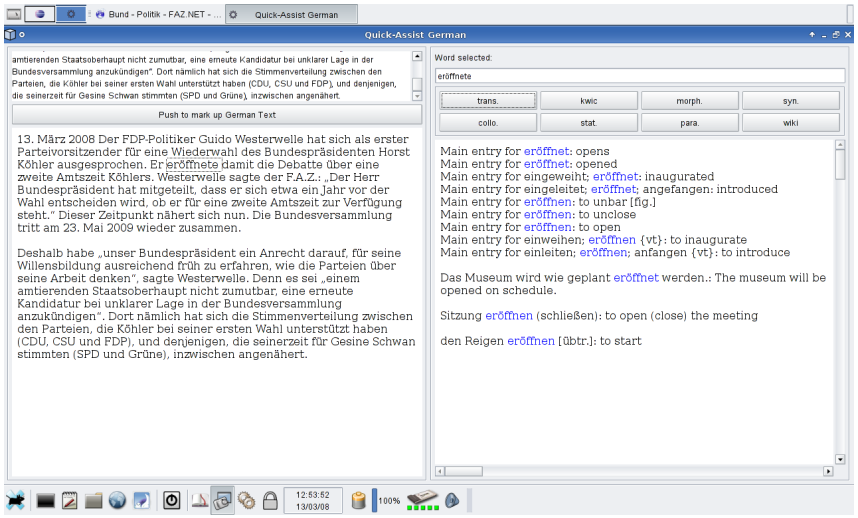

Figure 2. Concordancer output

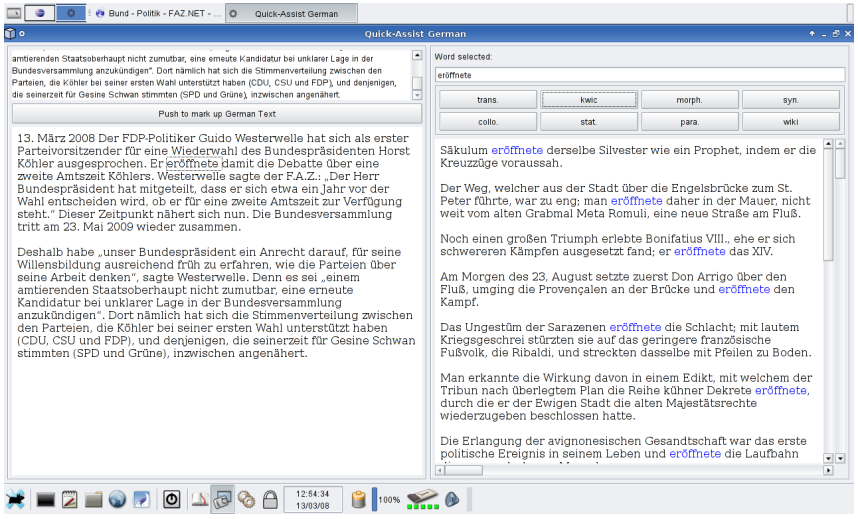

Figure 3. Query to German Wikipedia

An initial user study found that all four participants were able to use the application without any initial training to work effectively on a task that involved answering comprehension questions on a German text, which contained a number of low frequency vocabulary items, complex compounds, and the names of persons and institutions that the students were not familiar with. Two of the participants also reported after four weeks that they had used the tool successfully for the completion of assignments in their German courses or even for individual research.

One of the more interesting results of this small scale study, I find, is that modern learners are obviously able to learn to work with a piece of software quickly and efficiently. None of my subjects reported that they found the tool too complicated to use. They discovered a number of ways in which the software could be improved and were able to clearly identify the capabilities as well as the limitations of the program. It remains to be shown, but it can be hypothesised

that the exposure to electronic media, the familiarity with the Internet and the fact that the computer has become part of everyday life, has had the effect that most of us have developed strategies to filter information, assess the quality of sources, the suitability of resources for a specific task, and others. The age of the users also seems to create a difference in the use of strategies. In the user walkthroughs that were designed in a similar fashion to the ones outlined in Hémard (1999), there were two students in their early twenties, a retired teacher, and a professor emeritus. While both mature learners were trying to read and comprehend the text in a linear fashion, the younger learners used skimming techniques and were also somewhat more selective with the use of the program's functionalities and hence able to answer a greater number of the comprehension questions. A discussion of the user study comprising a detailed description of the learner walkthroughs and the results of a software evaluation by three German instructors is available in Wood (2010a).

After the completion of the initial study, the application was completely rewritten in C++ using the Wt libraries. It is still using the Wortschatz corpus, direct links to the German Wikipedia and canoonet (a lexical and grammatical German dictionary). Instead of using a wordform/baseform list to determine the baseform for dictionary look-ups, the new version uses TreeTagger, which was developed by the IMS Stuttgart. The new version of QuickAssist is web-based and can be used by the general public. The source code of the application is released under the GPL Version 3.0 and will be provided to interested CALL developers upon request.

5. Conclusion

The conclusions that can be drawn from this discussion are that tutorial ICALL software has its place in language teaching. For the time being, it appears that its use will remain confined to the beginning and lower intermediate learners. In order to foster learner independence, CALL software, and other learning software could include content that aims at the development of critical thinking, learning strategies and other skills that make a successful independent learner.

It appears that learners are able to handle a considerable amount of responsibility when it comes to the use of software tools. My subjects displayed an initial tendency to "play around" with features of the program (they would, for example, try to look up function words in Wikipedia). This seems to serve the purpose of getting acquainted with the technology and to probe its limits. After this phase, however, they started to become more and more selective in the use of features and developed strategies to quickly complete individual comprehension tasks.

Instructors are encouraged to give their students the chance to work with NLP tools and learn how to use them for their learning process. All of my subjects expressed excitement about the fact that these tools enabled them to read authentic German texts that were more difficult than what they would normally expect to be reading and that they could do so fairly quickly. While online dictionaries are widely used now by students, few students are aware of what information they can find in a corpus or other NLP tools like morphological analysers.

References

Bailey, S. M. (2008). *Content assessment in computer-aided language learning: meaning error diagnosis for English as a second language.* Unpublished doctoral dissertation. Ohio State University.

Boulton, A. (2010). Data-driven learning: taking the computer out of the equation. *Language Learning*, *60*(3), 534-572. doi:10.1111/j.1467-9922.2010.00566.x

Colpaert, J. (2004). *Design of online interactive language courseware: conceptualization, specification and prototyping. Research into the impact of linguistic-didactic functionality on software architecture.* Unpublished doctoral dissertation. University of Antwerp, Antwerp.

Dokter, D., Nerbonne, J., Schürcks-Grozeva, L., & Smit, P. (1998). Glosser-RuG: a user study. In S. Jager, J. Nerbonne, & A. Van Essen (Eds.), *Language teaching and language technology* (pp. 167-176). Lisse, The Netherlands: Swets & Zeitlinger.

Heift, T. (1998). *Designed intelligence: a language teacher model.* Unpublished doctoral dissertation. Simon Fraser University, Burnaby.

Heift, T., & Schulze, M. (2007). *Errors and intelligence in computer-assisted language learning; parsers and pedagogues.* New York: Routledge.

Hémard, D. (1999). A methodology for designing student-centred hypermedia CALL. In R. Debski & M. Levy (Eds.), *Worldcall: global perspectives on computer-assisted language learning* (pp. 215-228). Lisse, The Netherlands: Swets & Zeitlinger.

Higgins, J. (1988). *Language, learners, and computers: human intelligence and artificial unintelligence.* London: Longman.

Holland, V. M. (1995). Introduction: the case of intelligent CALL. In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), *Intelligent language tutors: theory shaping technology* (pp. vii-xvi). Mahwah, NJ: Lawrence Erlbaum.

Jager, S., Nerbonne, J., & Van Essen, A. (Eds.). (1998). *Language teaching and language technology.* Lisse, The Netherlands: Swets & Zeitlinger.

Johns, T. (1991). Should you be persuaded - Two samples of data-driven language materials. *ELR Journal*, *4*, 1-16. Retrieved from http://goo.gl/iOMI9

Julien, R., Makarova, V., & Wood, P. (in prep). *How are modern languages departments coping with budget cuts in Canada.*

Koller, T. (2007). *Design, development, implementation and evaluation of a plurilingual ICALL system for Romance languages aimed at advanced learners*. Unpublished doctoral dissertation. Dublin City University, Dublin.

Levy, M. (1997). *Computer assisted language learning: context and conceptualization.* Oxford: Oxford University Press.

Nagata, N. (1992). *A study of the effectiveness of intelligent CALI as an application of natural language processing.* Unpublished doctoral dissertation. Universtiy of Pittsburgh, Pittsburgh.

Nerbonne, J. (2003). Computer-assisted language learning and natural language processing. In R. Mitkov (Ed.), *Handbook of computational linguistics* (pp. 670-698). Oxford: Oxford University Press.

Nunan, D. (1997). Designing and adapting materials that encourage learner autonomy. In P. Benson & P. Voller (Eds.), *Autonomy and independence in language learning* (pp. 192-203). London: Longman.

Oxford, R. L. (2008). Hero with a thousand faces: learner autonomy, learning strategies and learning tactics in independent language learning. In S. Hurd & T. Lewis (Eds.), *Language learning strategies in independent settings* (pp. 41-66). Bristol: Multilingual Matters.

Pennycook, A. (1997). Cultural alternatives and autonomy. In P. Benson & P. Voller (Eds.), *Autonomy and independence in language learning* (pp. 35-53). London: Longman.

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching*. New York: Cambridge University Press.

Schmenk, B. (2006). CALL, self-access and learner autonomy: a linear process from heteronomy to autonomy. In T. Harden, A. Witte, & D. Köhler (Eds.), *The concept of progression in the teaching and learning of foreign languages* (pp. 75-90). Bern: Peter Lang.

Schulze, M. (2001). *Textana. Grammar and grammar checking in parser-based CALL*. Unpublished doctoral dissertation. University of Manchester, Institute of Science and Technology, Manchester.

White, C. (2008). Language learning strategies in independent language learning: an overview. In S. Hurd & T. Lewis (Eds.), *Language learning strategies in independent settings* (pp. 3-24). Bristol: Multilingual Matters.

Wood, P. (2010a). *QuickAssist. Extensive reading for learners of German using CALL technologies*. Unpublished doctoral dissertation. University of Waterloo, Department of Germanic and Slavic Studies, Waterloo. Retrieved from http://hdl.handle.net/10012/5680

Wood, P. (2010b). Review of Transparent Language system complete edition (German). *Calico Journal*, *28*(1), 229-237. Retrieved from https://calico.org/memberBrowse.php?action=article&id=839

Websites

Canoonet: http://canoo.net

General Public Licence (GPL): http://www.gnu.org/licenses/gpl-3.0.html

QuickAssist: http://linguistics.usask.ca/quickassist/German.wt

TreeTagger: http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger

Wikipedia: http://www.wikipedia.org

Wortschatz project: http://wortschatz.uni-leipzig.de

Published by Research-publishing.net, not-for-profit association
Dublin, Ireland; Voillans, France, info@research-publishing.net

© 2011 by Research-publishing.net

Second Language Teaching and Learning with Technology: Views of Emergent Researchers
Edited by Sylvie Thouësny and Linda Bradley

Rights: This article is published under the Attribution-NonCommercial -NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Under this licence, the contents are freely available online as PDF files (http://dx.doi.org/10.14705/rpnet.2011.9781908416001) for anybody to read, download, copy, and redistribute provided that the author(s), editorial team, and publisher are properly cited. Commercial use and derivative works are, however, not permitted.

Disclaimer: Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it was not under consideration for publication elsewhere. While the information in this book are believed to be true and accurate on the date of its going to press, neither the editorial team, nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

Trademark notice: product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Copyrighted material: every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net
Cover design by © Raphaël Savina (raphael@savina.net)

ISBN13: 978-1-908416-01-8 (Paperback - Print on demand, black and white)
Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

ISBN13: 978-1-908416-00-1 (Ebook, PDF, colour)

British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.