INTERACTIVE MULTIMEDIA APPLICATION FOR ONLINE LANGUAGE LEARNING

Veronica SCURTU[1], Vasile BUZULOIU[2]

În această lucrare propunem o nouă soluţie interactivă, personalizată şi motivantă de a învăţa o limbă străină în timp ce urmăriţi emisiunea favorită în limba originală pe un dispozitiv mobil sau pe laptop. Scopul este crearea unei noi forme de conţinut audio-vizual interactiv - video cu chestionar lingvistic - care va estompa diferenţa dintre formatul academic de modă veche şi aşteptările noii generaţii, propunând conţinuturi foarte atrăgatoare şi interactive în scopul învăţării unei limbi străine.

In this paper we propose a new interactive, individualized and motivating solution to learn a foreign language while watching and enjoying one's favorite original language video on a mobile device or a laptop. We are creating a new form of interactive audiovisual content – the linguistic quiz-enabled video – that will fill the gap between old-schools academic formats and the expectations of the new digital generation, by proposing highly interactive and attractive audiovisual contents for language learning purpose.

Keywords: multimedia content; interactivity; data processing; data language acquisition (SLA)

1. Introduction

The digitalization of content and the development of internet based networks have accelerated the media content production and distribution and led to an increased individualization ("one man, one device") and personalization ("to my taste") of media consumption.

The media consumption has become more individualized and customized due to the development of digital media and of more efficient portable devices. In this respect the rapid emergence and massive democratization of nomadic devices (mobile phones: 4 billion devices on the planet already, and personal computers: in a family, several individual laptops are typically replacing the usual single family desktop computer) has also supported this trend.

The access to knowledge is now more democratic and universal than ever. At home or on the move, at your own pace and according to your own schedule,

---

[1] R&D Engineer, Telecom SudParis, Evry, France, email: scurtu.veronica@gmail.com
[2] Prof. Applied Electronics and Informatics Engineering Department, University POLITEHNICA of Bucharest, Romania

you can easily download a tutorial to learn how to garden, play golf, or use any kind of software. The young generation has fast interactive digital lives which they want to translate as much as possible into their real ones. This digital easiness is an opportunity for education when specific educational content is created, a trend currently known under the generic name of edutainment. A domain where this trend is even clearer is foreign-language learning.

Among other fields, linguistic learning may benefit greatly from this trend if new types of content are created, specifically for the requirements of this new generation of learners, suitable for individual usage, portable, customized, flexible, interactive and entertaining.

In this paper we propose a solution that has its roots in the Edutainment (Education and entertainment) field and consists in automatically generating language comprehension quizzes based on foreign-language audiovisual content (TV series, documentaries, news). The proposed approach turns any foreign language video into a true interactive teaching material, by combining different fields of expertise, such as video analysis and meta-data extraction, information search and automatic language processing.

The remainder of this paper is structured as follows: in section 2 we point out the related work done in the field of online learning, presenting some of the most well-known and used online applications. Section 3 presents our contributions to the field of Multimedia Assisted Language Learning. In section 4 we described in more detail the system architecture. Section 5 concludes with an outlook to future work.

2. Language Acquisition through Multimedia Content

In language acquisition theories, the classic communicative approach claims that learning emerges through language production, i.e. a focus on speech and writing. On the other hand, a more recent school of thoughts is advocating the comprehension approach [1] referring to several methodologies of language learning that emphasise understanding of language rather than speaking.

Byrnes [2] proposed to develop listening comprehension in priority over speaking as it is the first natural step of language learning. Another study [3] shows that by increasing the exposure to authentic oral contents, such as videos, films and television programs, the listening comprehension skills are improving.

Since the main reasons European citizens point out for not studying foreign languages are lack of time and motivation, there is a need for a flexible, motivating and non-time consuming method [4]. Second language (L2) learning is facilitated because the young learner is already familiar with the topic or the content in the first language (L1) context. Meaningful learning takes place when the new material to be learned is related to something already known [5].

A 2007 Japanese study [6] on the use of Nintendo DS "English training" game (1 million units sold in Japan) showed improved results for junior high school students in English vocabulary.

We built our system on the hypothesis that the combination of the interactivity introduced by a quiz with the video game-like scoring system will lead to regular practice. Younger users, in particular, will see it as a form of entertainment, rather than education. Mobile devices (mobile phones, iPod Touch, PSPs, UMPC or laptop PCs) are the perfect platform for such service.

Built on a content-based approach and enhancing the development of listening comprehension competencies, our solution meets the expectations of the digital generation, introducing a new appealing and interactive type of learning oriented audiovisual content.

The demand for online learning is growing all over the world. Language learning institutions everywhere are waking up to the opportunities and benefits being offered by a fully-integrated online learning environment.
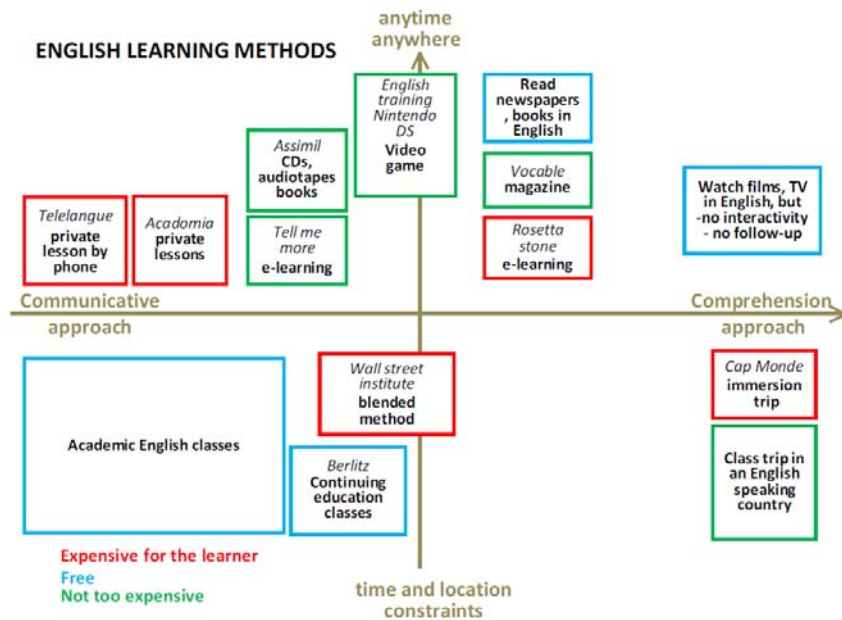
Fig. 1. English Learning methods

Although online comprehension has been a widely studied area for researchers in first language (L1) reading for quite some time, it has not been extensively explored in L2 research.

There are a several online tools that allow you to learn or improve a foreign language. Next, we will present the most common used for learning English as a second language.

English 360 [7] is a platform of blended learning with LMS (Learning Management System) and authoring tools created by Cambridge, covering levels A1 to C1 of English and American English, and its target public is the higher education and continuing education. It lets teachers create custom courses to be delivered in class, online or blended. There are no synchronous chat rooms or video conferencing tools. The platform offers online components but is not meant to substitute face-to-face sessions.

A similar e-learning platform is the English Campus [8] from Macmillan that offers authoring tools for higher and continuing education students who wish to learn American or British English, with a focus on general and business language and on exam preparation, combining face-to-face teaching and online support materials. The plus of this platform is that from some content it uses also video content and it allows voice recording. Students, teachers and administrators can upload their documents, audio files, images and presentations directly on their English Campus. The 'My Files' functionality gives teachers the option to introduce their own materials into the Campus and allows them to share the content they uploaded with their classes or to keep it private. At the same time students can send their homework to their teachers, who can then choose to share it with the rest of the class. The exchange and personalization opportunities are endless.

Digital Publishing [9] proposes a platform that has also integrated a voice recognition system. The application is addressed to self-study higher and continuing education students who wish to learn American English, German, Spanish, French or Italian, with a focus on general and business language and on exam preparation, with facilities for iPod and mp3 players. Every unit has a variety of exercises. You always have grammar explanations in the right top corner of the screen and you can right-click on the words and get the translation, or listen to the pronunciation of the word by different native speakers. You can correct the answers as many times as you like. If you don't know how to answer you can choose to see the solution.

MyEnglishLabs [10] is a new component from Macmillan that allows the teachers and the students to study part of the course online. Performance charts allow teachers to monitor performance at a glance and identify class or individual weaknesses which require further work.

The advantages and disadvantages of these e-learning platforms are summarized in Table 1.

Table 1

Online platforms

Name	English 360	English Campus MEC	Digital Publishing	Practice Online MPO	MyEnglish labs	Longman English Interactive	Cambridge Financial English
Editor	Cambridge	Macmillan	Digital Publishing	Macmillan	Pearson	Pearson	Cambridge
Video	×	●	×	×	×	✓	✓
Voice recognition	×	×	✓	×	×	×	×
Voice recording	×	✓	✓	✓	✓	✓	✓
Interface language	×	×	✓	×	×	✓	×
Program personalization	✓	✓	✓	×	✓	✓	✓
Authoring tool	✓	●	×	×	●	×	×
Learning Management System	✓	✓	✓	×	✓	✓	✓
Integrated Blended resources	✓	✓	×	✓	✓	✓	✓

● = only for some resources

3. Contributions

In the current paper, we propose a new interactive and motivating method to learn a foreign language while watching one's favourite original language video on a portable device. For example, on your way home you watch your favourite TV show that you podcasted on your mobile, and once every 3 to 5 minutes, your level of understanding is verified through a quiz that appears on the screen asking questions about the fragment that has just been watched.

Our solution integrates state-of-the-art technologies that rely on multi-source content aggregation (web, video), ontology definition, indexation, data extraction, meta-data processing, webcrawling, as well as state-of-the-art findings from research in the performance of educational systems in knowledge acquisition.

Three types of relevant audiovisual contents in English are used, taking into consideration a large number of parameters, including the video content's

duration and format, the different ontology types, the existence of related web knowledge databases: news, TV series and documentary.

A base of generic meta-questions was created after the analysis of the video contents, among which vocabulary, oral comprehension and general knowledge questions. Because one of our mid-term goals is the development of an application that can be used both on PC and mobile device platforms, only text multiple choice questions will be used.

A major objective of our work is reaching high level information through the use of low level visual/audio information found in video content. The audio-visual analysis of the content will allow detecting and retrieving objects of interest that will be employed to elaborate the quizzes.

The question and the correct answer are generated using the objects of interest that are detected, from the simplest (keywords) to the more complex (images), which also serve as query examples for searching in remote databases. Moreover, a certain number of false answers are also necessary. For this purpose, we develop a tool that processes the data received from the video and searches for new information in external databases using a technique similar to Webcrawling on Wikipedia and ontological resources such as Wordnet and WordReference.

While the text analysis allows the creation of a wider range of questions, the visual descriptors can be used only to generate "Who", "Where" or "What" questions. For example, when we want to identify a person performing an action, we use face detection and recognition. The scene change detection algorithm is employed when it is necessary to find out where location of the action (inside/outside). Gesture detection can also be used in order to test the user's attention to the visual content. For general knowledge questions, colour descriptors and/or basic object recognition are used (household objects, logos etc.).

After the extraction of the visual descriptors, our authoring tool proposes a question and 4 answers (1 correct and 3 false ones) that can be afterwards reinserted into the video content. On the user side a player is executed which contains an MPEG-4 content within which the interactive quiz is inserted. The video is periodically stopped and a set of questions about the scene that has just been viewed is displayed on the screen (Fig. 2).

Fig. 2. Preview of the interactive quiz inserted into the video

4. System architecture

The system architecture is presented in the figure below (Fig. 3). This system was created in order to offer non-technical users the possibility to modify and/or improve an online multimedia content. The result of the content modification operations is attractive, interactive and easy to use for the end-user. To visualize it, both a PC and a mobile terminal can be used.

The application consists in developing a software engine that will automatically generate a quiz on the content of a video (for example TV series, news, and documentaries), quiz which will be then reinserted into the video. For example, you watch the BBC news on your mobile phone, iPad or PSP3 in the subway, and every 3-5 minutes a quiz with questions from the news you have just watched is displayed to verify your level of comprehension.

The first implementation step consisted in creating a few web pages in an already existing web platform[3] that are used for uploading the multimedia content on the server, displaying it and editing the content. For their conception, we have used the web application framework Django written in Python. One of the main strengths of this platform is the model-template-view architecture that allows the parallel development of independent applications and their quite easy integration, just by changing certain settings of the project.
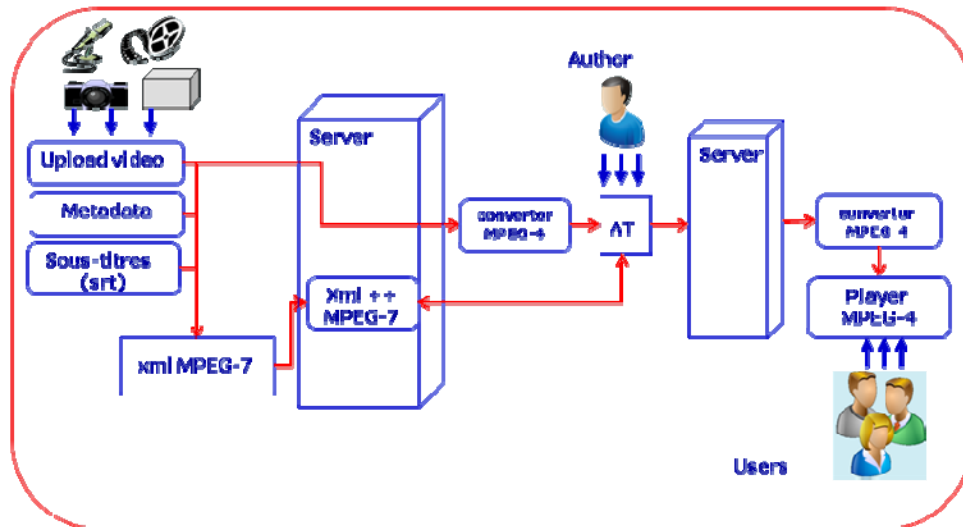
---

[3] www.mymultimediaworld.com

Fig. 3. Overview of the system architecture

The entire workflow of the defined system is detailed below.

4.1.   Upload

A user uploads on the Internet page a video content and the associated subtitle file. To send multimedia content, the user uses a web form where the file is sent by the server utilizing the POST method of the HTML language. Starting from a form, the user enters several pieces of information describing the content: the media type, the media title, a short description, keywords, if he want to make it public or not etc.

If the video file is not in MPEG-4 format, the video is first encoded in this format and then send to the MySQL database. If not, the video is pushed directly in the database.

4.2.   Data Processing

At this moment a plug-in is executed which creates an XML MPEG-7 file containing metadata and the subtitle representation in MPEG-7 standard. After generating the XML file, the API responsible for the analysis of the subtitles' text and for the generation of the quiz is executed. The output of this API is an "enriched" XML file. It is virtually the same as in the previous step, with the quiz part automatically added. The quiz generation application used is created by the French enterprise Jamespot[4] and it is based on the Stanford Parser [11].

---

[4] http://www.jamespot.com/

4.3.   Authoring

This step consists in reading and interpreting the XML file containing both the subtitles and the quiz. It content is displayed in the authoring page (Fig. 4), and certain information can be modified by the user.

Our approach makes use of an open API that has integrated a plug-in in charge with the text analysis and generation of questions. The authoring tool interface and its functionalities are presented in detail in the next section.

4.4.   Video Enrichment

Once all the changes have been made and the XML file has been re-generated, it is sent to another plug-in generating a BIFS (Binary Format for Scene) file. BIFS is a binary format that represents a description language for MPEG-4 scenes and they have the exrension .bt (BIFS Textual).

The main advantage in using BIFS and the MPEG-4 format is the possibility of creating an attractive and interactive content, allowing the user to interact and even modify the multimedia content.

The retrieval of this data from the APIs allows the creation of the interactive scene. The last step is the generation of the MPEG-4 interactive video file corresponding to the .bt file created at the previous step.

4.5.   Data Consumption

Once the multimedia content has been through the defined workflow, it can then be packaged for delivery. The video files created can be visualized both locally using an open-source MPEG-4 player, Osmo4[5], and remotely (server-side) using the Osmozilla plug-in.

5. Authoring Tool

The left side of the table where the subtitle lines are displayed cannot be modified; it exists only to offer the user a complete vision on the questions and their correlation with the lines. As it has already been mentioned, the questions will be generated automatically based on the lines. The right side of the table is fully dynamic.

In the bottom-left side you can see the subtitles that are separated in to sub-groups. Each sub-group has a number of questions associated to it.
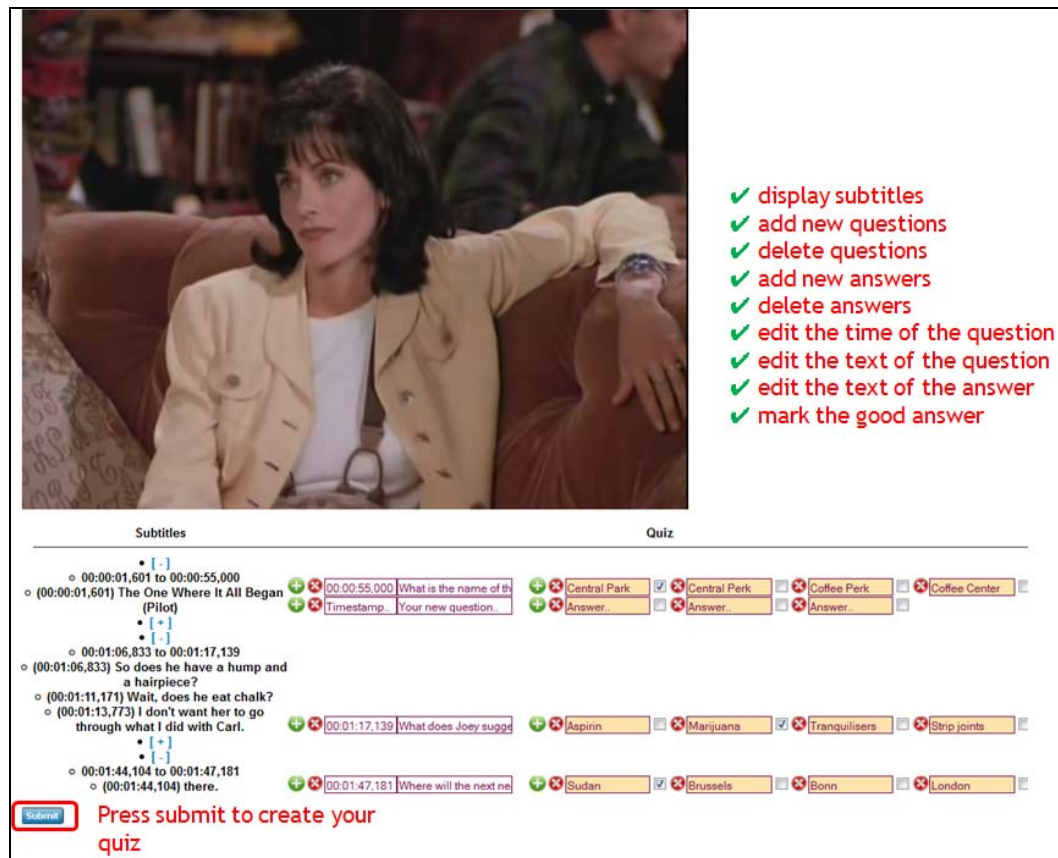
In the authoring tool you have different permissions:

- For adding a new question inside a sub-group you have to press the '+' button that is in front of one of the questions from that sub-group;

---

[5] http://gpac.wp.mines-telecom.fr/player/

- For deleting a question, you just have to press the 'x' button that is in front of it;
- For adding a new answer to an existing question, you have to press the '+' button that is in front of the answers;
- For deleting an answer, you have to press the 'x' button that is in front of it;
- You can also edit the following fields: time stamp of the question, question's text, and answers' text;
- For each question you have to mark what is the right answer, by ticking the box after it;

Fig. 4. Snapshot of the online authoring tool

The reason why each question can be deleted or modified is simple: besides the fact that the table will include initially automatically generated questions and possible answers (so probably not all the questions will be correct), the user has the freedom to completely modify the multimedia content.

The next step offers the user the possibility to choose one of the two quiz visualization interfaces (Fig 5).
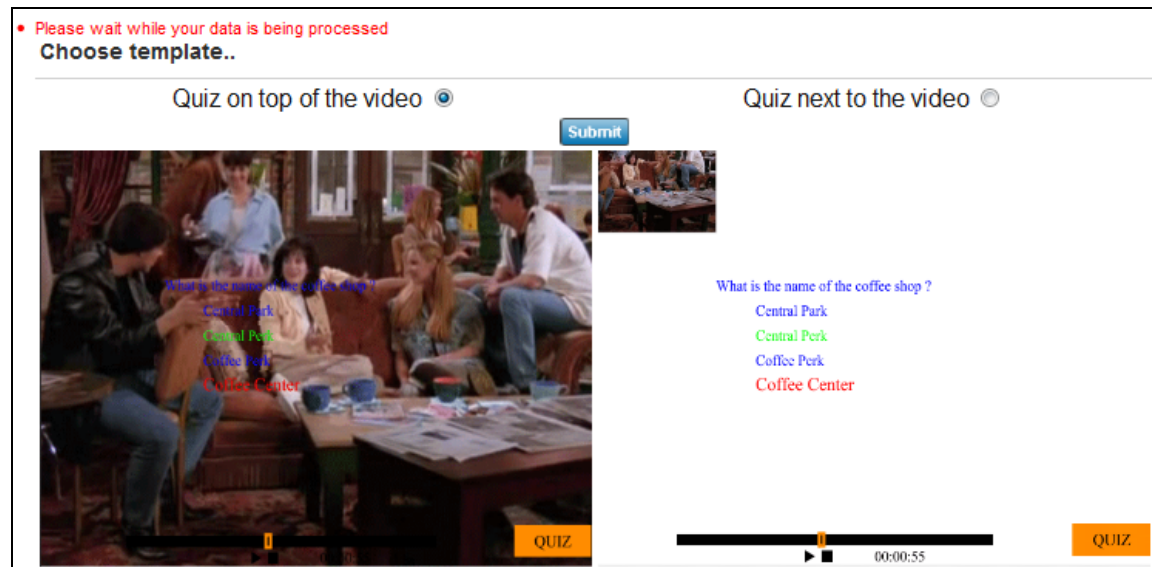

Fig. 5 Overview of the quiz interfaces

The data transmitted by the previous page are still held back; they will be used only when the user chooses the pattern and clicks the "Submit" button. At this moment the questions, the answer variants and the correct answer (also selected by the user checking one of the checkboxes) will be used to create a new .bt file corresponding to the pattern the user has selected.

Although the proposed framework can be fully automatic and self-sufficient from the technical point of view, given the complexity of the task of creating pertinent linguistic content, human intervention is required in order to ensure the relevance of the quiz elements.

6. Conclusions

In this paper, a new interactive, individualised and engaging solution to learn a foreign language while watching and enjoying one's favourite video on a laptop or a mobile device was introduced. First, we presented a general overview of our contributions, describing the types of questions and the methods used to generate them and their answers. Then we described in more detail the system architecture of the online platform we have created.

In the language learning context, TV series are the best material for testing the potential of visual descriptors extraction, as the characters and objects are recurring in the long run. We have used images from the TV series "Friends"

because it mirrors the type of program the young generation is mainly interested in and is most likely to watch.

This research is still in progress and the perspectives for our research work are focused on the one hand on the enrichment of the platform, by creating new APIs for other types of descriptors (e.g. face and object recognition, color, text interpretation, action recognition). This will lead to the possibility of generating new types of questions, but more importantly they will enrich even more the MPEG-4 video scene. On the other hand the learning content can be improved, by adding new functionalities. One of the possible enhancements is the use of interactive subtitles. This could allow the user to click on any word from the subtitles in order to obtained more information about it (e.g. definition, translation, synonyms etc.).

In our future work we will organize two different user trials. The first one will assess the authoring tool and the panel will be composed of English teachers, while the second will consist in test carried out on a larger number of students (>50) in order to verify the effectiveness of the proposed system in foreign language learning.

R E F E R E N C E S

[1] *S.D. Krashen*, "Exploration in Language Acquisition and Use: The Taipei Lectrues," 2003, Portsmouth, NH:Heinemann.
[2] *H. Byrnes,* "The role of listening comprehension: A theoretical base." 1984 Foreign Language Annals, 17, 317-329.
[3] *C.A. Herron and I. Seay, "*The effect of authentic oral texts on student listening comprehension in the foreign language classroom", 1991, Foreign Language Annals, 24, 487-495.
[4] *Eurobarometer "* European and their languages *",* February 2006.
[5] *D.P. Ausubel, J.D. Novak and H. Hanesian,* "Educational psychology: A cognitive view (2nd edition)", 1978, New York: Holt, Rinehart and Winston.
[6] *M. Casamassina*. "Nintendo Sales Update". IGN, July 2007.
[7] "English 360 – Commercial and pedagogical benefits of a Web 2.0 based blended learning platform", Cambridge University Printing House, 2010.
[8] Macmillan English Campus, http://www.macmillanenglishcampus.com/
[9]Speexx – The perfect way to lean a language, Digital Publishing, http://www.digitalpublishing.fr/
[10] "My English Lab – An online component to complement your favourite English language course from Pearson", Pearson Longman Education, 2009 http://www.pearsonlongman.com/ae/myenglishlab/
[11] *M.C. de Marneffe, B. MacCartney and C. D. Manning*, "Generating Typed Dependency Parses from Phrase Structure Parses", 2006,. In LREC 2006.