

# EU 4 U: An educational platform for the cultural heritage of the EU

**Maria Stambolieva**

Centre for Computational and Applied Linguistics  
New Bulgarian University  
Sofia, Bulgaria  
mstambolieva@nbu.bg

## Abstract

The paper presents an ongoing project of the NBU Laboratory for Language Technology aiming to create a multilingual, CEFR-graded electronic didactic resource for online learning, centered on the history and cultural heritage of the EU (e-EULearn). The resource is developed within the e-Platform of the NBU Laboratory for Language Technology and re-uses the rich corpus of educational material created at the Laboratory for the needs of NBU program modules, distance and blended learning language courses and other projects. Focus being not just on foreign language tuition, but above all on people, places and events in the history and culture of the EU member states, the annotation modules of the e-Platform have been accordingly extended. Current and upcoming activities are directed at: 1/ enriching the English corpus of didactic materials on EU history and culture, 2/ translating the texts into (the) other official EU languages and aligning the translations with the English texts; 3/ developing new test modules. In the process of developing this resource, a database on key figures and important places, objects and events in the cultural history of the EU will be created. The target users of the resource are students aged 8-12. Along with creating a useful teaching resource for local students, the project also addresses problems facing migrant and minority learners, with the ambition to contribute to the social insertion of young learners in an age of increased people mobility. It has the support of the UN Refugee Agency (UNHCR).

## 1 Context

Whether young learners come to the EU from other parts of the world or move within the EU, they need to gain basic information about both their new country of residence and the European Union. Some of this information is taught or otherwise acquired by native children at a very early

age, and is often taken for granted. Acquaintance with the rich and varied cultural history of Europe and the EU states can be an important factor for stimulating the desire of young learners to become part of this culture. The development of such thematic content in several EU languages, coupled with graded language tuition and translation, and via an open education multilingual e-platform, can offer substantial support to both students of foreign languages and cultures and to young learners accompanying their parents in transnational mobility.

Eurostat 2015 data indicate that as a result of intra-EU mobility and immigration from third countries, EU societies are becoming increasingly diverse. Roughly 10 per cent of the population of the European Union were not born in their country of residence; an increasingly large percentage of this population are children under 15. These “migrant children” have, as a rule, low educational performance and tend to leave school early. In spite of the considerable attention this problem has received in the past decades — from the Council of the EU 1977 Directive 77/486/EEC on the education of children of migrant workers to the more recent EC Green Paper “Migration and mobility: challenges and opportunities for EU educational systems”, these children continue to suffer from “a negative penalty associated with migratory status”.

The integration of migrant children is becoming a key problem for the successful development of the EU, for which — as analysts point out — solutions are yet to be sought, found and implemented. (Algan et al., 2010, p. 25) report that education systems do make efforts to integrate immigrant children, “though it is much harder to say whether progress is as fast as it could be”. (Dumcius et al., 2013, p. 5) stress the necessity to adopt an integrated approach to the problem of child migrant

inclusion, combining linguistic and academic support, parental and community involvement and intercultural education — an aspect also stressed by (Nusche, 2009) and (, ed.).

Driven by the seriousness of the problems outlined, our objective is the development of a multilingual EU History & Culture learner pack — e-EULearn — for young learners, with a common educational content, both language and content-based tests and quizzes, as an accessible introduction to EU cultures and languages. We identify the following subtasks:

- combining foreign language learning with a EU cultural context;
- creating a supportive environment for personalised and individualised online tuition, also suitable for refugee, migrant and minority children in the language of their new country of residence;
- offering methodological support to schools and teachers with up-to-date resources, tools and reference materials.

## 2 Online Education and the NBU e-Platform

*An educational platform for the cultural heritage of the EU* was initiated by the team of the NBU Applied Foreign Languages undergraduate program, where students from 16 countries study together, and of the NBU Language Technology Lab, which has recently worked on the generation of educational content for the language and cultural integration of migrant and ethnic minority children. The project stems from our experience in addressing problems that face migrant and minority learners and draws on positive results: a/ the success of courses developed and taught by our program team for a multicultural audience of undergraduate students, as e.g. *The Languages of the EU*, *EU Cultural History*, *Electronic Resources of the EU*, and others; b/ the success of distance learning language courses supported by an integrated platform for text annotation, semi-automatic extraction of training content and automatic assessment. The project builds on this know-how and on the accumulated corpus of texts on the history and culture of the EU to develop didactic content suitable for a target age group of students aged 8-12 and by extending the functionalities of the software support.

The e-EULearn pack for young learners is an educational resource presenting EU history and culture in graded CEFR texts and drills. The design of e-EULearn is based on the methodological principles of e-learning. Most often used as a supplement to traditional classroom tuition, e-learning is an invaluable means of increasing the overall effectiveness of the process of teaching — especially if sufficiently well planned and conceived as an integral part of this process. Effective distance or blended learning requires no less careful planning and preparation than traditional brick-and-mortar classes; and the simple addition of available online videos or tests to existing educational content might lend a course additional flavour but will not necessarily increase its effectiveness. While the needs of trainees are symmetrical, the groups are often heterogeneous. In designing the e-EULearn pack, we follow (McDonough and Shaw, 1993) who define a well-designed e-course as one that provides for personalisation and individualisation of the learning process. In line with (Hemingway, 1986) we will insist on providing students with the relative freedom to choose from a rich learning content, supplying them with the appropriate reference materials and with sufficient training tasks.

Online (distance, blended or self-directed) education is a learning option which, with the opportunities it creates for individualisation and personalisation of the educational process, is well suited to the needs of learners coming from different linguistic and cultural backgrounds and/or learners in need of additional tuition. An important condition for effective online learning is the availability of electronic resources and supporting software. NBU is among the leading HEIs in the region in the field of both e-education and language technology. Our accumulated know-how can now be put to the benefit of school education by directing efforts towards the development of tools and resources suited/adapted to the needs of our target learners.

## 3 The NBU e-Platform

The NBU e-Platform for Language Teaching and Research is a modular, versatile tool designed to provide 1/ course development support for native and foreign language (and literature) teachers and lecturers, 2/ data and tools for corpus-driven and corpus-based lexicography, corpus and contrastive

linguistics, 3/ an environment for research, experimentation and comparison of new methods of language data preprocessing. It integrates: 1/ an environment for creating, organising and maintaining electronic text archives and extracting text corpora: a repository of domain-specific texts, further classified in accordance with the Common European Framework of Reference for Languages (CEFR); 2/ modules for linguistic analysis, including a lemmatiser, an in-depth POS analyser; a term analyser; a syntactic analyser; an analyser of multiple word units (MWU — including complex terms, analytical forms, phraseological units); a parallel text aligner; a concordancer; 3/ a linguistic database allowing linguistic analysis to be performed on either a single text or a corpus of texts, with options for corpus manipulation (reduction in size or expanding with additional texts from the archive) without loss of information; 4/ modules for the generation and editing of vocabulary or grammar drills; 5/ modules for the extraction of linguistic information directly from texts/corpora or from the data base. The environment for the maintenance of the electronic text archive organises a variety of metadata which can, individually or in different combinations, form the basis for the extraction of text corpora. Following linguistic annotation, secondary (“virtual”) corpora can be extracted. The platform with its data base can thus be used to support a variety of linguistic activities — from the generation of drills to the compilation of corpus-driven and corpus-based reference materials — glossaries, thesauri, dictionaries or grammars. The e-Platform can support all EU languages; of these, only three have been made active so far: English, French and Bulgarian.

The architecture of the system is modular and consists of input modules, modules for preprocessing, processing, analysis and data storage, and output modules. The input modules provide user interface for different linguistic tasks and work independently. The architecture allows for the addition of new modules, as well as various modules for automatic (pre)processing. Each of the preprocessing modules can be implemented independently and added to the system at an appropriate stage. The text-based drills generator can export the teaching content to the educational platform of the user. In the case of New Bulgarian University, the export format is Moodle XML format for Moodle Quiz Module. The e-platform feeds

the question banks of Moodle with three types of question items: Fill in the Blanks, Matching, and Reordering, each with a number of subtypes. The architecture of the system is maintainable and extendable.<sup>1</sup>

For the purposes of this project, the NBU platform is being developed in several directions:

- its functionalities are extended to allow work with new languages: French, German, Spanish, Russian;
- the annotation tags are tailored to the new languages of the project;
- new annotation tags are added in answer to the focus on EU history and culture;
- new test options will be added to the test-generating functionality.

#### 4 Virtual Corpora and the Re-use of Didactic Material

The “Virtual corpora” function of the e-Platform allows the extraction of token-based or annotation-based corpora from existing corpora and files. Figure 1 presents a list of subcorpora, mostly generated from Wikipedia and tourist guides.

A virtual corpus can contain a word form (e.g. the 3<sup>rd</sup> person sg, pronoun *it*), a lemma, a part of speech (e.g. Numeral), the value of a grammatical category (e.g. Perfective Aspect — for Bulgarian or Russian). For the purpose of the EU 4 U project, the following new subcategories were introduced in the annotation module: Noun/Proper/Person, Noun/Proper/Place, Noun/Common/Hist&Cult (e.g. *tumulus* or *artifact*), Numeral/Cardinal (1885), Numeral/Ordinal (15<sup>th</sup>), Adjective/Proper (Byzantine), Adjective/Hist&Cult. This latter category contains adjectives identified as forming indicative contexts

<sup>1</sup> Cf.: Stambolieva et al. 2017a. M. Stambolieva, M. Hadjikoteva, M. Neykova, V. Ivanova, M. Raykova. 2017. The NBU E-Platform in Teaching Foreign Languages for Specific Purposes. Proceedings of the 13th Annual International Conference on Computer Science and Education in Computer Science, Albena. Stambolieva 2017b M. Stambolieva, M. Hadjikoteva, M. Neikova, V. Ivanova, M. Raikova. 2017. Language Technologies in Teaching Bulgarian at Primary and Secondary School Level: the NBU Platform for Language teaching. Proceedings of RANLP 2017 Workshop” LTDHCSEE: Language Technology for Digital Humanities in Central and (South-) Eastern Europe, 32-8. ISBN 978-954-452-046-5



Figure 1: Virtual corpora.

for phrases referring to historical and cultural heritage. An indicative list of verbal contexts was similarly compiled.

Virtual corpora can, like other corpora in the platform, be manipulated: split or merged (united) with other corpora — Cf. Figure 2.



Figure 2: Union of corpora.

The unification of virtual corpora facilitates the generation of the didactic materials and tests focused on EU culture and history — by providing generalized lists of important people/places/events. Together with the text files and text-based corpora, they form the basis for the generation of our training materials — quizzes and tests.

## 5 Deriving EU History and Culture Tests from Virtual Corpora

The test generator of the e-Platform has two main modules: Matching, Fill in the Blanks and Re-ordering. Only the first two modules are actively used for this project. For them both, test generation starts by a choice of corpus, then of the test

segments — relevant segments, or all segments (Select All) on which the test will be based — Cf. Figure 3.

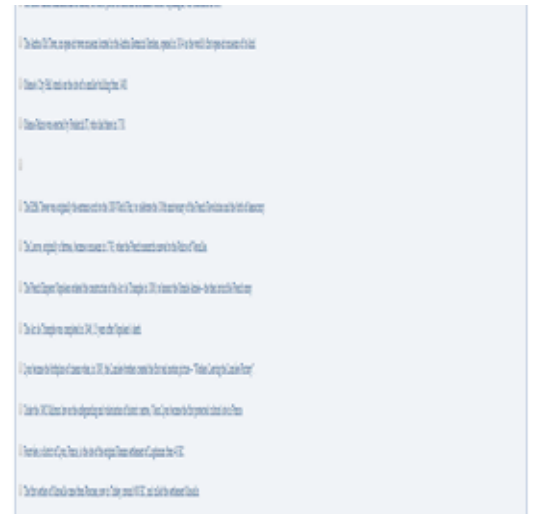


Figure 3: A CEFR A1 level corpus for four EU countries.

Texts and corpora, whether virtual or not, should be annotated (manually or otherwise) prior to test generation.

For the Fill in the Blanks module, tokens must be lemmatised and POS-tagged. The POS-tagger of the platform has been extended with the new subcategorisation values.

For the Matching module, the “definition” slot of the POS tagger can be used in a number of ways to link people to biographical notes, places and events, places to people and events, events to dates, people and places. The slot can also be filled with photographs and pictures of people and places.

When the selected segments are uploaded in the test generating module, the relevant tag must be chosen, the number of occurrences indicated (each, every second, etc.), and the desired test type selected — multiple choice (in either drag-and-drop or dropdown format) or gapfill (open cloze). Once the test has been generated, the platform allows post-editing.

All texts entering the e-Platform are graded for CEFR levels A1-C2. For the purpose of this project, our corpora consist of short A1 and A2 level texts. Levels are defined with the help of Laurence Anthony’s AntWordProfiler<sup>2</sup> but a more fine-grained and CEFR-level based profiler

<sup>2</sup><https://www.laurenceanthony.net/software/antwordprofiler/>

of texts is planned as a project-related task. Another task yet to be defined and implemented is the manner of integration of the aligned bilingual texts into the EULearn lesson pack, the types of reference and training resources that can be derived from them.

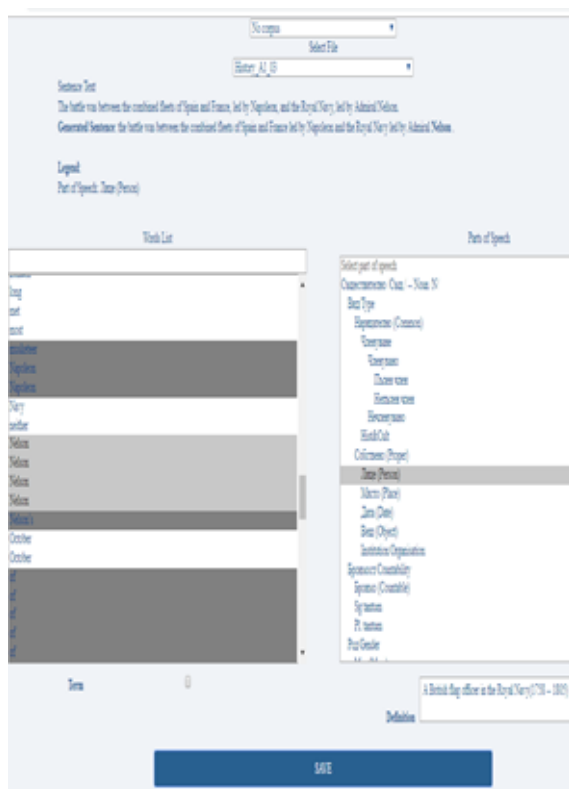


Figure 4: Annotation.

Texts and training materials will be carefully sequenced in Moodle lessons to facilitate step-by-step learning: students will first be asked to link people and events to centuries, then do multiple choice training tests, followed by open cloze ones. They will learn to link names and places with photographs, pictures or short texts, to link people, places and events with countries. The Moodle educational platform where the training exercises are exported also allows the addition of audio/video tips to the tests and quizzes, and learning there can be fun. To ensure personalization, a considerable choice of didactic materials and audiovisual aids as well as a choice of languages will be offered. The possibility for self-paced learning in the flexible Moodle lesson structure offers individualization of the learning process.



Figure 5: Generating a test on the history and culture of the EU.

## 6 Conclusion

The texts, data base and training materials developed within the project will form a useful resource with data related to people, places, objects and events in the cultural history of the EU, and links between them. While initially planned to provide learning support for children migrating across borders, the EULearn pack can offer useful additional materials for native children learning a foreign language and culture or for other educational purposes. The data base and the aligned texts can offer support for translation studies and linguistic research.

## References

- Yann Algan, Christian Dustmann, Albrecht Glitz, and Alan Manning. 2010. [The economic situation of first and second-generation immigrants in france, germany and the united kingdom](#). *The Economic Journal*, 120(542):F4–F30.
- Rimantas Dumcius, Hanna Siarova, Jana Huttova Ides Nicaise, and Indre Balcaite. 2013. Study on educational support for newly arrived migrant children.

- Julia Szalai (ed.). 2011. Contested issues of social inclusion through education in multiethnic communities across Europe. *EDUMIGROM Final Studies*.
- P. Hemingway. 1986. Teaching a mixed-level class. In *Practical English Teaching*, pages 18–20.
- J. McDonough and C. Shaw. 1993. *Materials and Methods in ELT*. Blackwell, Cambridge, Mass.
- Deborah Nusche. 2009. [What works in migrant education? A Review of Evidence and Policy Options.](#) *OECD Education Working Papers*, (22).