A PLATFORM FOR LANGUAGE TEACHING AND RESEARCH (PLT&R)

Maria Stambolieva

mstambolieva@nbu.bg

Laboratory for Language Technologies

Valentina Ivanova   Mariyana Raykova

v.ivanova@nbu.bg   mraykova@nbu.bg

Digital Innovation Laboratory

NEW BULGARIAN UNIVERSITY, SOFIA, BULGARIA

PLT&R MAJOR APPLICATIONS

PLT&R is a platform for corpus creation and annotation with applications in:
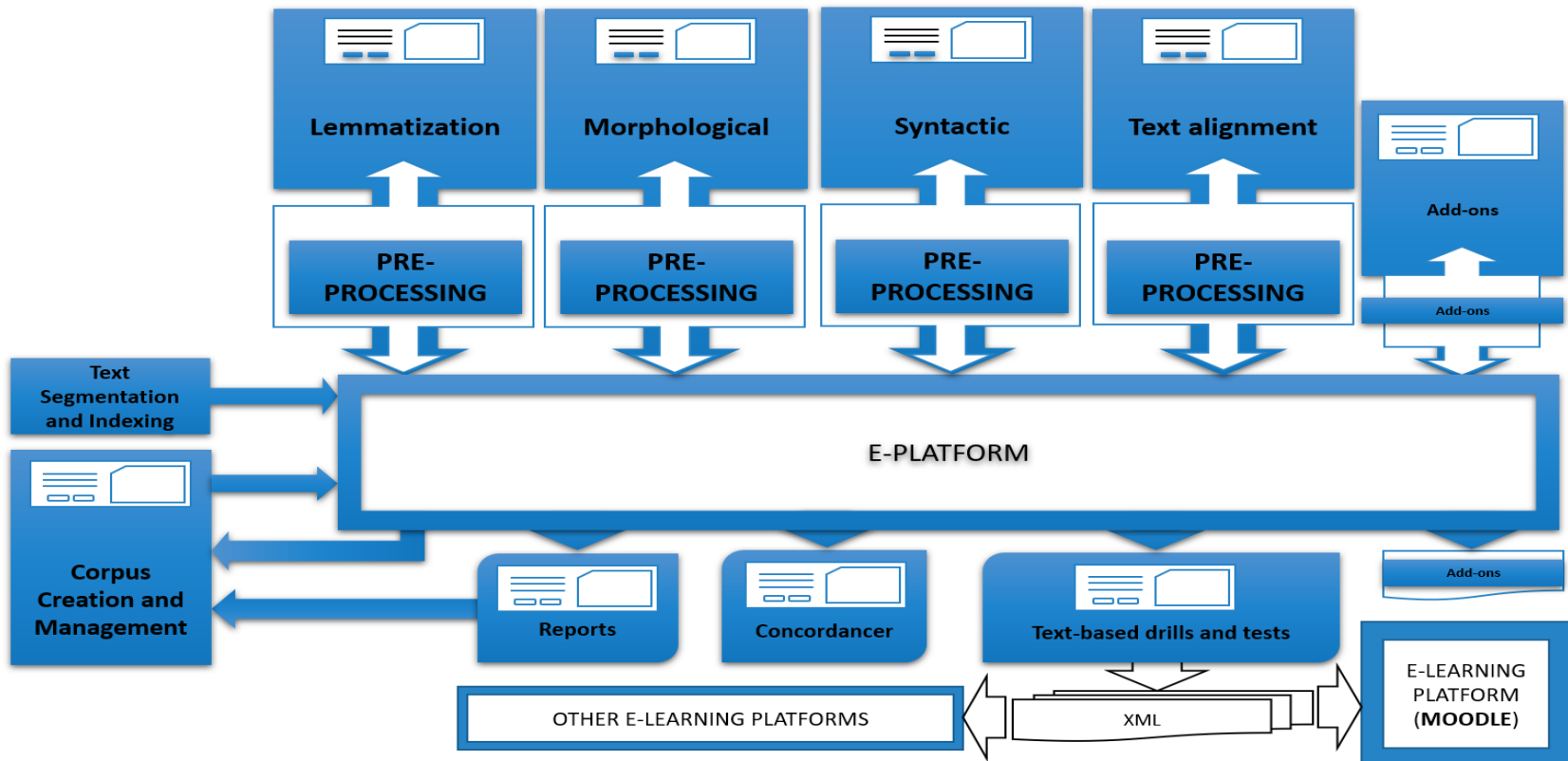
- language teaching

- translation

- research

CORPUS LIST

Corpus Name		Keywords
File Name		Language [Select option]
Corpus Description		Book/Part of Book ○ book ○ part of book ○ publication
Author [Select option]		File name
Gender ○ male ○ female ○ not known		Number of Pages
Status ○ author ○ coauthor ○ creator		Publisher
Date		Start Page
Date From		End Page
Date To		File Type ○ an empty new line is a paragraph ○ each new line is a paragraph
Domain [Select option]		File Level ○ not defined ○ A1 ○ A2 ○ B1 ○ B2 ○ C1 ○ C2

SEARCH CORPUS

THE PLT&R IS A MODULAR PLATFORM

- modules for corpus **creation** and **annotation**

- modules for the **generation of educational content**

CORPUS CREATION MODULES

text-based corpora ▶ virtual corpora

virtual corpora can be based on:

- **attributes:** e.g. author, year, CEFR and others;
- **annotation results:** POS analysis, syntactic analysis, combinations of these.

CORPUS ANNOTATION

The modules for corpus annotation are based on **entity-attribute associations**. Both entities and attributes are user-defined.

Entities can be: strings, morphemes, word forms, lexemes, phrases, syntactic functions, sentences

Attributes (e.g. morpheme types, parts of speech, grammatical categories and syntactic functions) are hierarchically organised and language specific. The set of attributes is extendable.

The platform also supports **alignment** at sentence level.

EDUCATIONAL CONTENT GENERATION

Types of educational content generated by the platform:

- drills: multiple choice, reordering, multiple matching, open the brackets, open cloze

- establishing translation equivalence

- training corpora

APPLICATIONS

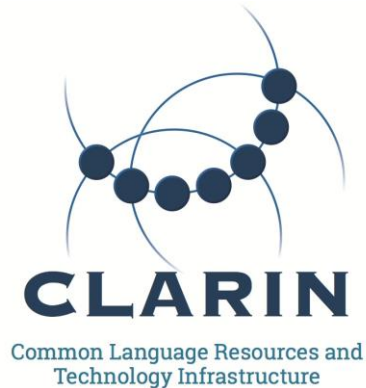Native and foreign language teaching

Corpus-driven and corpus-based lexicography

Corpus and contrastive linguistics

Translation

New methods of language preprocessing: research, experimentation and comparison of methods

PLT&R AND CLaDA-BG

The content organised and generated by the PLT&R can be integrated in the CLARIN part of the CLaDA-BG infrastructure.

PLT&R TEAM

Thank you for your attention!

Maria Stambolieva

mstambolieva@nbu.bg

Laboratory for Language Technologies

Valentina Ivanova

v.ivanova@nbu.bg

Mariyana Raykova

mraykova@nbu.bg

Digital Innovation Laboratory

NEW BULGARIAN UNIVERSITY, SOFIA, BULGARIA