

Deliverable 7.2 Draft of the platform structure (macrostructure)

1. Usage scenarios, relevant data types and display modes per target group

The usage scenarios have been derived from the results of the user study as well as from research on the use of learner corpora in language teaching and SLA. The following table provides an overview over concrete tasks that the prospective user groups might want to perform on the platform and includes a specification of relevant data types and metadata information as well as display modes.

Teachers

	aim	usage scenario / use case - task	data types	relevant information / features of learner language (LL)	relevant levels	display mode
1.	define / agree with colleagues on common rating criteria (also for testers)	extract a random sample of written tests on a specific tasks for sample texts by different metadata as e.g. L1, task type, CEFR level (of the takes / rated CEFR level), [special feature: list of sample texts per level → pre-defined]	unannotated learner productions	available metadata, esp. CEFR level of test, fair CEFR level	learner (ZH1/2)	entire text/ text section metadata
2.	prepare teaching material for students on different CEFR levels	search for frequent errors or learner language features from different linguistic fields (vocabulary, grammar, etc.) at different CEFR levels → by LL feature → by word / by lemma ... and display alternative/correct versions Target hypothesis (TH)	annotated learner productions	available metadata, esp. level of test and fair CEFR level features of LL: grammar, vocab., etc.	learner, ZH1, ZH2, ZH1Diff, EA1-2	LL-feature in context + target hypothesis (TH)
3.	explore typical errors in context	search for words, adjacent words, POS in learner texts and target hypotheses	annotated learner productions	POS, lemma, features of LL: grammar, vocab., etc.	learner, ZH1	LL-feature in context
4.	Underpinning the course schedule with lists of LL features specific for different CEFR levels & identify	extract feature list by CEFR level using different filter criteria as L1 and CEFR level (od test / fair CEFR level)	feature list	available metadata, esp. test level and ratings	learner, meta-data, EA1, EA2	feature list

	typical and relevant milestones/errors in learner language with reference to CEFR levels	[special feature: pre-fabricated lists]		linguistic annotations (statistical information – features per level)		
5.	use standardized samples for classroom presentation	→ see 1	unannotated learner productions	available metadata, esp. rating and test level	learner	entire text/ text section
6.	Use examples for classroom presentation	search for elements of text macro structure in learner productions (e.g. greetings, closings ...)	annotated learner productions	available metadata, esp. task types; annotations from the transcription	learner, EA2 greetings, closure	text section

Teacher trainers and testers

	aim	usage scenario / use case - task	data types	relevant information / features of LL	relevant levels	display mode
7.	create a common understanding of the CEFR levels and practice rating procedures of written texts	→ see 1				
8.	re-adjust teachers oversensitive to special L1 errors by having them re-rate MERLIN texts without showing the MERLIN ratings and then compare and discuss the differences / results	→ see 1 (5) and 6				
9.	Prepare teaching material for teacher training	search for learner texts by - features of LL and - meta-data	annotated learner productions	available metadata, esp. test level LL features	learner, ZH1, ZH2, ZH1Diff, EA1-2 meta-data	entire text / text section (LL-feature in context)
10.		creating sub-corpora & sampling texts by meta-data as e.g. L1, CEFR level, task type + institution	annotated learner productions	available metadata, esp. test level	learner, ZH1, ZH2, ZH1Diff, EA1-2	entire text/ text section metadata

				features of LL: grammar, vocab., etc.	meta-data	
--	--	--	--	---	-----------	--

Textbook authors

	aim	usage scenario / use case - task → <i>as for teachers / teacher trainers +</i>	data types	relevant information / features of LL	relevant levels	display mode
11.	Add usage notes to textbook materials, e.g. hints on correct use or suggestions to avoid the overuse or underuse of words or structures.	show error-prone categories for different CEFR levels	annotated learner productions	available metadata, esp. test level LL features	learner, ZH1, ZH2, ZH1Diff, EA1-2 metadata	statistical information on chosen LL features

SLA Researcher / Linguists

	aim	usage scenario / use case - task → <i>all of the above mentioned +</i>	data types	relevant information / features of LL	relevant levels	display mode
12	e.g. understand what types of errors occur in which contexts; which errors occur more frequently; understand what errors are typical for a certain L1;	search for a particular error category in different contexts, event. combined search for error categories + POS etc.	annotated learner productions	meta-data, features of LL (all levels) POS type of target lang. modification	learner, ZH1, ZH2, ZH1Diff, EA1-2 metadata	Key word in adjustable context frequency lists & distribution of chosen LL-features
13		show feature lists and frequency information for different error categories				
14		explore distribution of an error category over the different CEFR levels				
15		search for words, adjacent words, POS in learner texts and target hypotheses to explore typical errors using: - regular expressions - complex query syntax - sub-corpora (learner's L1, task type, LL-				

		features, ...)				
16	Understand what are differences among different groups of learners, e.g. for different L1	Compare feature lists of different groups of learners and find differences → <i>see also 4</i>	Annotated learner productions	meta-data, features of LL (all levels); possibly statistics on the occurrences	Learner, EA1-2 Meta-data	Parallel display of features lists
17	analyse the relationship between CEFR descriptors (e.g. “content jumps”) and learner language					

2. Macrostructure – Information pages

2.1 General aims and requirements

- design an interface that offers all search functionalities and satisfies the requirements of usability
- present search functionalities in non-corpus-linguistics style, but task-oriented
- address the specific needs of the different target groups
- present the project to the public
- give access to public project results and to the project documentation: final publication, user instructions, annotation scheme, annotation guidelines

2.2 Requirements on the design of the user interface

Standards that should be considered

- e-learning standards according to:** QPL - Qualitätsplattform Lernen: Teil A Qualität von Bildungsangeboten: http://issuu.com/delan/docs/qpl_qualitaetsplattform_lernen_teil_a
 - clear-cut target groups
 - target group orientation in terms of content and didactics
 - self-descriptiveness: clear navigation, help functions, glossary
 - technical requirements → browser-compatibility
 - data protection and privacy statement (if we go for saving user specific information in profiles / registration)
 - usability standards
- usability standards according to: ISO 9241** (standard covering ergonomics of human-computer interaction and dealing with principles that apply to the interface design)
 - suitability for the task → **minimal interaction**
 - self-descriptiveness → **clear indication for the user of his current location in the site, which options are available**
 - suitability for learning → **help is available: FAQ section; guided tour/ tutorial**
 - **controllability of the interaction** → trigger search, go back to search interface from every point
 - conformity with user expectations,
 - suitability for individualization,
 - error tolerance

Results from user studies on usability of corpora

- Findings from an end user evaluation of the search interface for a Spanish learner oral corpus: Campillos Llanos, L. (2012) Designing a search interface for a Spanish learner spoken corpus: the end-user's evaluation, in: Proceedings of the Eight International Conference on Language Resources and Evaluation LREC'12
- User wish to include
 - a diagram of the error taxonomy (with examples)
 - glossary of terms (with examples)
 - error frequency count per group of students with the same L1
 - further explanation of the aims of the tool and the type of user
- other recommendations:
 - teaching guide to be used along with the computer interface
 - a more dynamic menu design to simplify the page where the search criteria are selected

2.3 Site structure

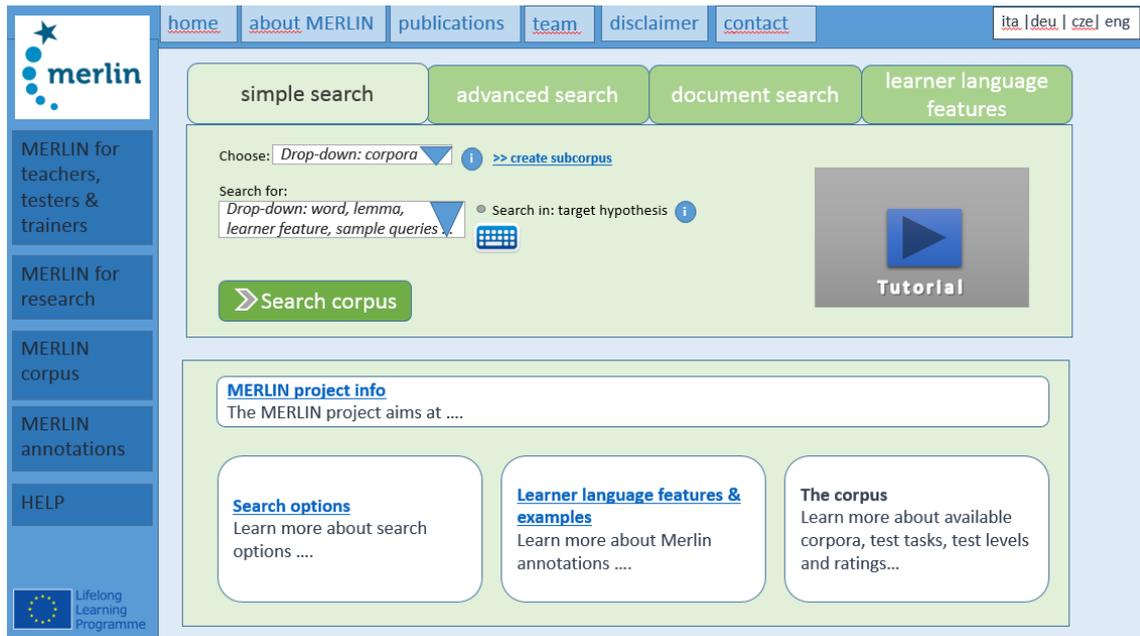


Figure 1 - Draft of platform macrostructure

Navigation

- **Top navigation bar:** provides quick access to more general content about the project: background and aims (about MERLIN), who is behind MERLIN (team), how can I get in touch with the team (contact)
- **Left navigation bar:**
 - categories presented here are:
 - information tailored to specific target groups & main usage scenarios of the site
 - information backing the search process
 - structure of pages: content can be structured a) as grouped set of collapsible panels or b) by providing sub-pages (possibility of introducing lower-level navigation)
- **Top2 - Search navigation bar:** Tabs to switch between the search options (for details see deliverable 7.3)

Homepage

- starting point for all search functionalities
- accessible from every sub-page
- info boxes → feature basic questions that user could be interested in when coming here for the first time (panels refer to one of the pages in the left navigation bar)
- offers several short-cuts to help functions and background reading

Page: MERLIN corpus

... provides all users with GENERAL information about:

- the corpus preparation and compilation: texts and tests they were extracted from; number of texts per CEFR level and language; Information on re-ratings and the MERLIN rating grid;
- available metadata: information about the author (e.g. L1, sex); information about the tests
- description of the test tasks

Page: MERLIN annotations

... provides all users with GENERAL information about the annotation process and the annotation levels:

- *General remarks on annotated learner language features and their sources* → annotation scheme
- structure of annotations in MERLIN / annotation layers explained using one example (e.g. *in a tabular format*)
- *General remarks on the concept of the target hypothesis*

2.4 User specific content

Page: MERLIN for teachers, testers & teacher trainers

- working scenarios for teachers, testers and teacher trainers
- with concrete use cases and reference to example searches / search pages

Page: MERLIN for researchers

- DETAILED information about:
 - the annotation structure / annotation guidelines
 - corpus architecture
 - NLP tools / machine learning components
 - including references to literature
- working scenarios for researchers

3. Supporting structures / user guidance

3.1 Tutorial

- should contain information on search functionalities – best opportunity to present information on routines in a very dense and concise way

3.2 Context sensitive help

- available in all search interfaces
- explains function of buttons / search functionalities in the search interface via tool tip

3.3 Help page

- offers help in a key word format → glossary
- is structured FAQ-like
- available from every point in the interface

3.4 Use cases and sample searches

- use cases and example searches are presented on target group pages
- example queries should be available for advanced search options