

WP 4 – User modelling

Deliverable 3.1 - Report on user relevance

Part 2 – Functional and technical aspects of the MERLIN platform

1 Report on technical requirements to the MERLIN platform

1.1 Introduction

The work carried out in WP3 user modeling aims at collecting users' requirements with respect to the functional and technical realization of the MERLIN platform. It helps shaping the technical setup and design of the tool and the underlying data, and thus ensuring the user relevance of the product. By means of an online questionnaire and individual expert interviews information on the situational and technical working context of users and on how users want to enquiry and access learner texts via the MERLIN platform is gathered.

The questionnaire (see attachment 2.5) is divided into four sections:

1. Functionality of the platform
2. Technical working environment
3. Working context
4. User profile

Section 1 focusses on the users' informational needs concerning learner texts and information attached to it. Furthermore, it investigates the users' preferences for data management and display. Section 2 examines technical variables of the users' work environment, including internet access, technical support, operation system etc. as well as familiarity with formal search specifications. Section 3 inquires the users' working context in terms of language use, team work and static vs. changing working locations. The questionnaire closes with section 4 on the users' profiles. This part corresponds to the user profiling in part 1 – linguistic aspects of the user study.

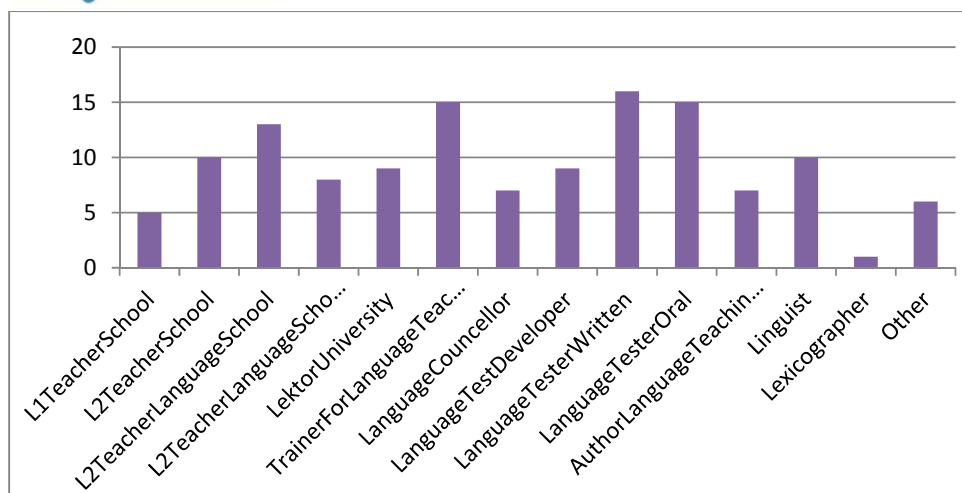


Figure 1: Professions of respondents

The questionnaire was distributed in all three project languages (Czech, German and Italian). Overall, 55 questionnaires were compiled. The respondents represent all target groups distributed as shown in Figure 1¹. The distribution of working languages is shown in Figure 2².

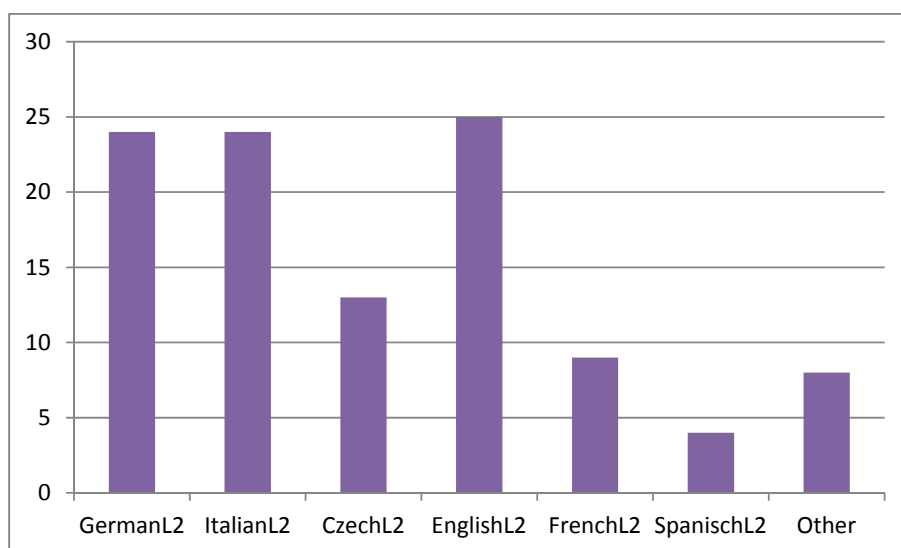


Figure 2: Working languages of respondents

Expert interviews were carried out with an expert for test development and written language testing and an expert for language teaching and testing, teacher training and consulting for autonomous language learning. Working languages of the experts are German and German/Italian. Also an expert in language

¹ Respondents were allowed to select more than one profession.

² Respondents were allowed to select more than one working language. Other languages include Russian, Polish, Japanese, Portuguese and Greek.

teaching and testing for Czech was interviewed. The three interviews followed a semi-structured guideline (see attachment 2.6) that expands on several questions of the questionnaire.

1.2 Results of questionnaires³

1.2.1 Functionality of the platform

Features of learner language

Users rate vocabulary and grammar the linguistic features most relevant for search and analysis of learner texts, followed by text characteristics and sociolinguistic criteria/text type. Orthography and form of texts are considered least relevant with punctuation on the lowest end of the scale. Apart from orthography all features of learner language are considered relevant for text searches by more than half of the people. Comparing responses for the different target groups teachers are mostly interested in grammar and vocabulary; trainers have a stronger focus on text characteristics than other groups; linguists are interested in vocabulary frequencies and sociolinguistic aspects and text types, while trainers are not much interested in vocabulary frequencies and sociolinguistic aspects and text types apart from the usage of formal conventions.

Figure 3 shows that different aspects of each feature are rated differently. For example, the absence or presence of certain words by itself is considered less indicative than the adequate usage of words. As well the register of a learner's text is less relevant for text searches than the occurrence of formal expressions.

³ See attachment 2.7 for a summary report of the responses.

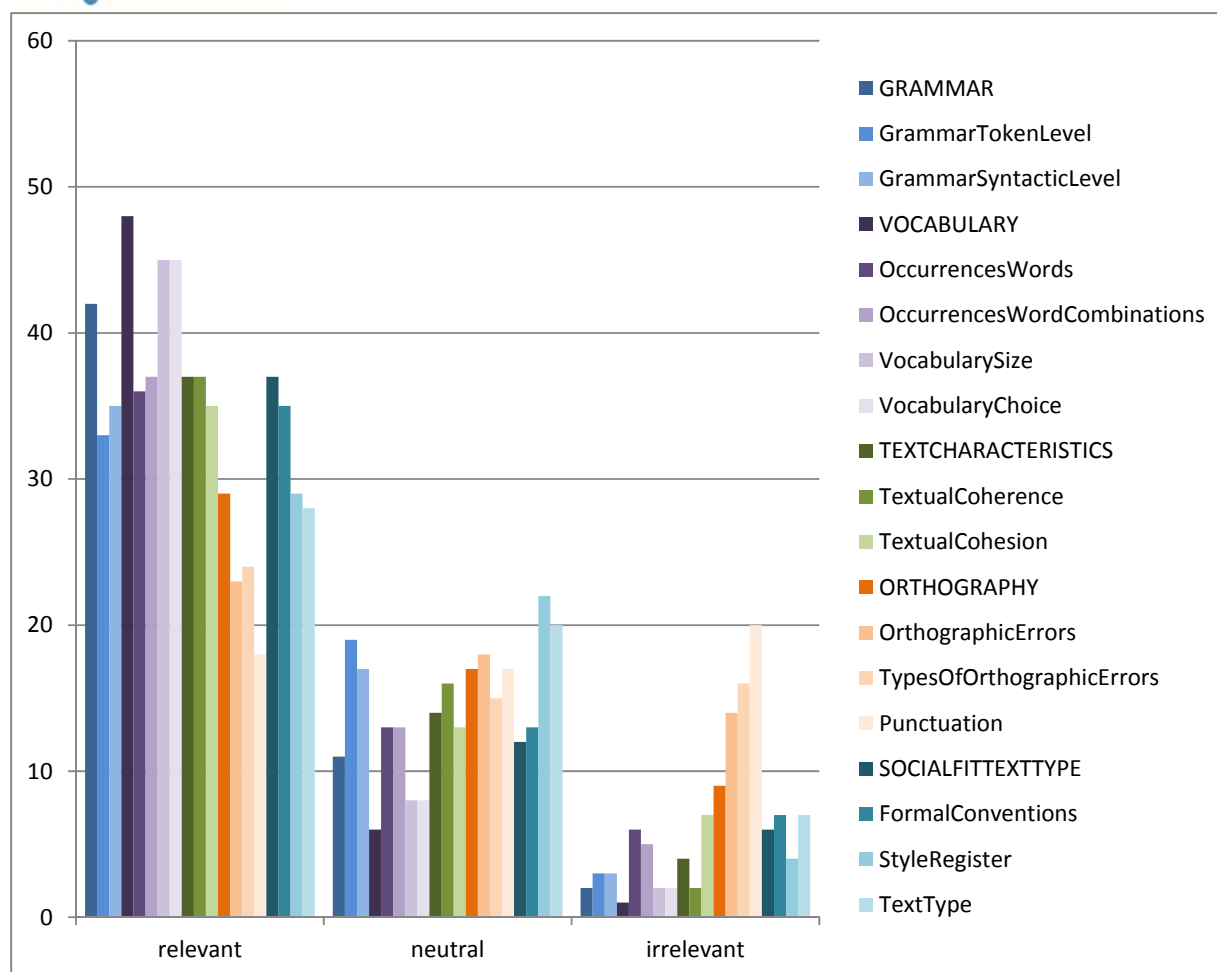


Figure 3: Features of learner language

Metadata

By the majority of users metadata are considered important for selecting and analyzing texts. Level and type of the test task are rated the most important type of meta information. Users also name learning context, aim of the test task and time for completing the task relevant. Comparing the responses of different user groups it turns out that metadata are most relevant for linguists and least relevant for teachers, with trainers and testers situated in between.

Filtering by quantitative data

Size of vocabulary and sentence complexity are the most important quantitative filter criteria, followed by the variety and frequency of learner language features (see Figure 4). In particular, frequencies of one syllable words and questions are mentioned by a Czech speaker, and collocations and word combinations by a speaker of Italian. Feature dispersion, text length and sentence length is given least importance. Interestingly sentence complexity and sentence length are found on different ends of the rating scale. As with metadata, quantitative filtering criteria are most relevant for linguists and least relevant for teachers, with trainers and testers in between.

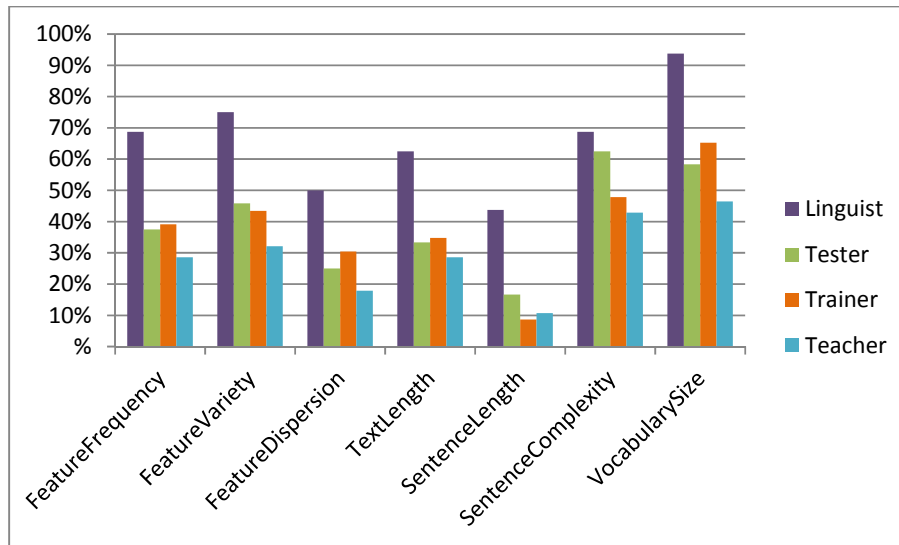


Figure 4: Quantitative criteria for text filtering

Size of unit for analysis

Groups of texts are the prior unit for analysis closely followed by single texts. Analyses on the level of sentences, paragraphs or text sections are less relevant.

Display context

Concerning the unit for displaying search results users agree in requiring a wider context of either the full text or a text section (see Figure 5 below). One respondent suggests to have a context window that can dynamically be expanded. Only for teachers a context of one sentence is more relevant than a larger section of the text, while they still rate a full text display the highest. The presentation of features without context is considered irrelevant by all user groups.

Presentation of learner language features

The majority of users votes for options for highlighting features of interest. Being able to inspect features in more detail in a separate window or an expandable text field on demand is desired only by half of the polled users.

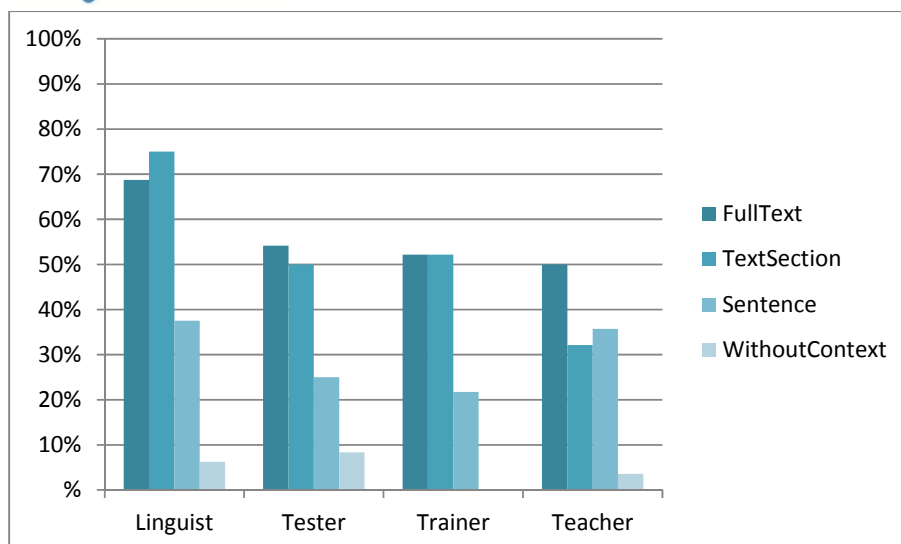


Figure 5: Text context for display of search results

Other functionalities

All of the suggestions for additional analysis and display functionalities are well received by users in the following order: listings of features and feature values are most relevant, followed by metadata and task texts, systematic comparison of learner texts and groups of learner texts and access to CEFR scales. We notice that listings of features are estimated very high as additional functionality, but very low when proposed as display contexts (see above). A confrontation of results by target group shows: for linguists feature lists and metadata are most important; by trainers and testers feature lists and test tasks are rated most relevant; teachers generally rate all types of additional information as less important than other groups do, with feature lists still the most relevant, but desired by less than half of the polled linguists.

Notes for texts

Being able to attach notes and annotations to texts and features is considered important by about 60 to 70% of the users. Among the three options, notes for feature types are least important.

Storage and export of data

Functionalities for storing information and exporting data are of high relevance for most users. In particular the availability of bookmarks and support for exporting texts and text sections are highly desired. Comparing replies by different user groups shows that for linguists, testers and trainers export functionalities are most relevant, closely followed by bookmarks. In contrast, teachers give most importance to bookmarking facilities and support for collaborative working. Among the prospective user groups, teachers are least interested in being able to export feature lists.

Usage of texts

Asked about what they would like to use texts for, the most frequent reply is preparing teaching material and using texts in the classroom for demonstration. At second position, users name the use of learner texts for assessment and as a mean to develop common assessment criteria with colleagues. To a lesser extent linguistic research analyses are mentioned.

1.2.2 Technical working environment

Internet access

Less than 10% of the users do not regularly have access to the internet in their working environment.

Internet speed

Around 10% of the users are using a slow or very slow internet connection.

Plugins and players for the web browser

More than 65% of the users need to get help with installing additions to their browsers or are at all unable to get additions.

Operating system

The wide majority of users is working in a Windows environment with about 10% of Linux and 10% of Mac OS users.

Web browser

Most users are working with Internet Explorer and Mozilla Firefox, followed by Chrome, Safari and Opera (see Figure 6 below).

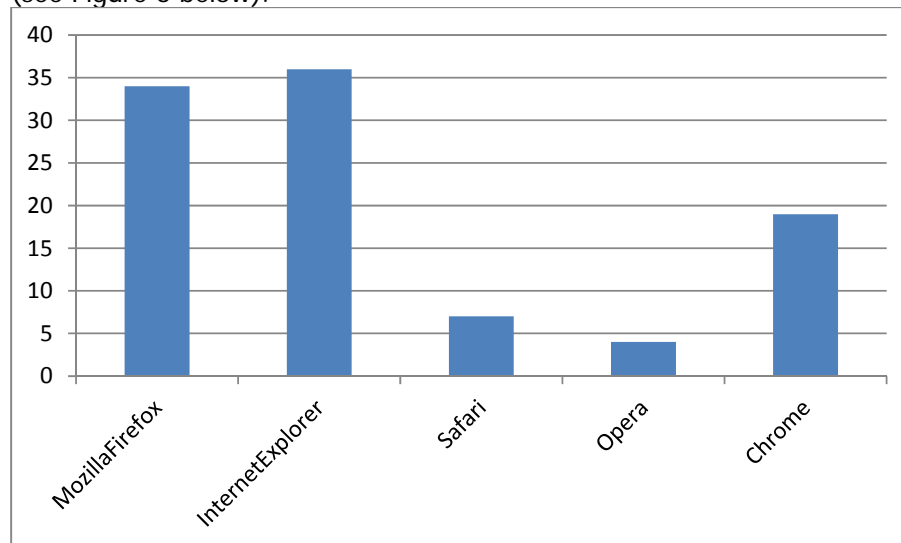


Figure 6: Used web browsers

Supported file types

More than 3/4 of the polled users expect problems in viewing .xml files, and up to 2/3 in viewing .txt or .odt files.

Familiarity with search specifications and annotation formats

The majority of users is familiar with lemma-based searches and regular expressions. Less than 20% of the users are familiar with reading XML data.

1.2.3 Working context

Workplace

The wide majority of users is occasionally working on different PCs.

Team work

More than 2/3 of the respondents are planning to use the MERLIN data together with colleagues or in the classroom.

Working languages

The wide majority of users is occasionally working with more than one of the project languages.

1.3 Results of interviews

1.3.1 Interview 1 - expert for test development and written language testing, German

In view of the prospective MERLIN platform, the expert's central concern is being able to compare competences of learners of a certain CEFR level as rated by MERLIN, with competences required by the test institution for that same CEFR level. That is, the user is interested to compare possibly different understandings of what belongs to a certain CEFR level, with the aim to better estimate learners' competence levels with respect to the framework. Providing CEFR ratings that are grounded on a profound theoretical and scientific elaboration, as done by the MERLIN consortium, is seen as a major strength of the MERLIN tool.

For accessing typical features for competence levels two types of functionalities are relevant: On the one hand the expert is interested in extracting lists of features for specified groups of learners (e.g. learners on level B1, with Russian as L1); on the other hand she is interested in looking up the CEFR level that a specified feature would belong to (e.g. CEFR level, in which formal address - German "Sie" - should to be handled well). Access to full learner productions is secondary. However, having texts with similar profiles (e.g. similar types and frequency of errors) grouped together is relevant. Also accessing prototypical texts for each CEFR level is relevant.

Functionalities for exporting feature lists are considered very important. Though not relevant for the expert herself, she considers it helpful to allow for flexible export modes, including or excluding the annotation of specified features on demand.

Filtering of the data is relevant on the content level (e.g. for linguistic features like 'communicative structure', 'register', 'grammar', 'orthography') as well as by metadata (primarily on CEFR level, L1 and age of the learner).

Concerning the infrastructure of the platform, contextualized help functions are preferred over video tutorials. A team area is not relevant for the expert, but is estimated as possibly beneficial for other users. The expert particularly expressed her interest in a feedback functionality for getting in touch with the MERLIN team on content issues.

Rating scales for the different test tasks should be published together with learner productions. Also examples of the vocabulary range for different CEFR levels are considered relevant to render the competence levels of the CEFR more transparent.

The potential use of the platform for learners is not clear to the expert.

1.3.2 Interview 2 - expert for language teaching and testing, teacher training and consulting for autonomous language learning, German and Italian

The expert envisions two primary use cases for working with the MERLIN platform: consulting in autonomous language learning, and training of teachers and language testers.

For consulting in autonomous language learning MERLIN texts would be used to illustrate the CEFR levels with the aim to help the learner estimate her current competence level, help her understand what competences are needed for the next level and based on that define learning targets. The expert states the following use case: The learner is asked to produce a text for one of the tasks that MERLIN texts are based on. In a follow-up step the consultant together with the learner is placing the learner's text among the learner texts provided by MERLIN. The texts are compared with respect to different linguistic features (e.g. means of verbal expressions, vocabulary, grammar, etc.), to give the user an orientation and rise her awareness to specific competences required for a certain competence level.

For training of teachers and testers, MERLIN texts would be used to illustrate the CEFR levels with the aim to arrive at a consensus on assessment criteria and teaching targets⁴, allow for comparability of assessments and rise the teachers' and testers' awareness of their rating behavior. The expert describes the following use case: MERLIN texts of different competence levels are handed over to teachers and testers for evaluation, without indicating the ratings assigned by MERLIN. Results of the ratings are discussed among the teachers and are compared to the ratings assigned by the MERLIN consortium. In the same way, features detected by teachers and testers are compared to the features annotated in MERLIN texts. By this trial evaluation and discussion teachers and testers are enabled to establish a common ground for their ratings and are encouraged to reflect on their rating behavior. Having available for reference the MERLIN ratings, which are done by a scientific working group, is seen as a particular benefit.

Both use cases require that the MERLIN platform makes available learner texts also without annotations and ratings added. Also, tasks that learner productions are based on should be accessible.

Metadata are relevant not only for filtering, but also for sampling data. For example, texts by learners with different L1s should be sampled to get a balanced set of learner language features and not introduce a bias towards certain feature types⁵. Being able to extracting groups of texts that are similar to the teachers'/testers' working realities is likely to increase motivation to use the tool. In addition, metadata on text type and content are required and considered particularly helpful for filtering data. Functionalities for filtering learners' texts based on linguistic features are relevant for illustrating what features correspond to which CEFR levels, create a transparency and rise self-awareness on the weighting of features in test assessment. Functionalities for extracting lists of features prototypical for certain competence levels are relevant for preparing teaching schedules for courses and setting learning targets for individual learners. Search and export functions should be coupled in the MERLIN tool, that is filtering based on metadata and text internal search for features of learner language (macro and micro structure), e.g. extracting lists of connectors for argumentative texts. It is expected that learners will most benefit from contextualized examples of text, while lists of features are most helpful for teachers. Export functionalities should be provided for all levels: full text documents, lists with features or sections of texts including certain features.

Video tutorials are the most welcome way to introduce the platforms functionalities, followed by contextualized help functions. A team area, including a forum for discussions among users as well as

⁴ Possibly targeted to a specific context, like e.g. courses for university students who have to attend lectures in a foreign language.

⁵ The expert mentions that teachers/testers sometimes tend to get over-sensitive to features of learner language that they are confronted with very often, no matter how serious the features really are.

discussions about decisions taken by MERLIN is rated as very helpful, even if it might not be well accepted by the less computerized users.

1.3.3 Interview 3 - expert for language teaching, testing and test development, Czech

Q 10 Clear criteria according to which the texts can be searched, possibility to search for more parameters at the same time, possibility to look for individual words to see collocations, count the frequency, etc.

Q 11 As mentioned above, more filters at the same time, e.g. to look for texts at A2 by Russian speaking students. Also to search for the text type (e.g. a letter).

Q 12 It depends. I like to vary things.

Q 13 I would not like it. If I need to find something, I can look it up. It would irritate me.

Q 14 Writing notes could be helpful, but is not necessary, she is not used to using it. It also depends on how it would look like, if the style would suit her.

Q 15 + 16 She has never worked together in this sense with her colleagues. Maybe at a different type of school/course. She cannot imagine it in during common lessons.

1.4. Summary of findings and indications for the platform design

The interviews confirm that access to concrete examples of rated learner texts are a relevant point of reference for language teaching and testing as well as for teacher training. The profound theoretical and scientific elaboration of CEFR ratings by the MERLIN consortium is very much appreciated.

Concerning the MERLIN platform, the survey and interviews show that the proposed ways to search and display the MERLIN learner data are relevant to the prospective users. Some differences in the preferences of the different user groups (divided into linguists, testers, trainers and teachers) can be observed from the questionnaires. Generally, there is a strong focus on the learners' texts in themselves. For the MERLIN platform it is thus important to present searched data in their wider (and maybe further adjustable) textual context, and groups of similar texts should be accessible together. The interviews show that contextualized examples are most needed in working with learners, while lists of features are rather relevant for teachers and in teacher training to understand correspondences of concrete features and CEFR levels and to define learning targets and course schedules. The interviews further show, that different modes for accessing the data should be offered: on the one hand lists of features should be extractable for specified groups of learner, on the other hand one should be able to look up on what CEFR level a specified feature is situated. In addition, functionalities for filtering by metadata and searching for features should be combinable. Also searching for words and collocations should be supported. Metadata is considered relevant not only for filtering, but also for sampling data (e.g. texts by learners with different L1s). Text typology as well as text content could be relevant additional types of metadata.

Consulted in the interview, one expert for language teaching and testing stated the need for accessing learner texts without annotations and ratings attached to it. As well, task descriptions, rating scales and details on CEFR features were considered relevant building blocks of the MERLIN platform. As for the display of primary data (and annotations and meta information attached to it) prospective users favor the highlighting within the text, over data presentation in separate windows and pop-ups.

The majority of users is occasionally or regularly working with more than one of the project languages in parallel. The MERLIN platform should therefore accommodate respective searching facilities and not keep data of the three languages strictly separated.

Mainly linguists show a strong interest also in information derived from the texts or attached to it, like statistics on word and feature frequencies and metadata. For trainers and testers access to extra textual information that defines the situational context of the text production is important, e.g. the underlying

test task. As well, they are interested in inspecting feature lists. For the MERLIN platform it is advisable to place functionalities that allow for the aggregation of information and the quantitative analysis in a separate area to text selection and search. This is indicated by the differing responses to what presentation formats users favor in their text searches (question 7, where respondents express little interest in feature lists without context) and what information they are interested in accessing in general (question 9, where users state a considerable interest in feature lists).

Generally, the MERLIN platform should target the presentation of all functionalities to the users informational aims instead of presenting corpus linguistics style query functionalities. This becomes apparent in the users' estimation of filtering criteria (see question 5) where sentence complexity is rated important by many users while sentence length is not – although indeed sentence length can be a relevant indicator for sentence complexity. Advanced searching modes employing regular expressions or distinguishing between word and lemma searches are familiar to about half of the polled users. When including these facilities into the MERLIN platform detailed instructions for their usage have to be foreseen. As well XML data cannot be assumed to be handled well by the majority of users but needs to be processed before being presented to the users.

Regarding the management of MERLIN data, the platform should support different types of data exports (ranging from full text to feature lists), bookmarking facilities and support for collaborative working, as is particularly requested by teachers.

The technical working environment of the respondents suggest that an offline version of the MERLIN platform is not required (most users do have regular access to the internet⁶), while a space for storing and sharing information is indicated, as a majority of users is regularly changing their workplace (e.g. office vs. classroom) and is likely to use data in collaboration with colleagues or students. The interviews confirm the potential use of a team area. In particular, the experts express an interest in being able to comment on the MERLIN data and possibly get in touch with the MERLIN consortium on content issues. Among prospective users mostly the Windows operating system and standard browsers like Internet Explorer and Mozilla Firefox are used. Requiring special plugins or players for being able to run the MERLIN platform should be avoided as a majority of users will not be able to use them without asking external support. As well, data exports are best provided in standard formats like .pdf or .doc, which are widely supported in the users' environments. Manuals for using the platform in terms of video tutorials or contextualized help functionalities are encouraged while extensive written documentation are less favored.

⁶ As mentioned above, the data collection via an online questionnaire introduces a bias to this point. We will take into consideration relevant experiences among institutions of the MERLIN partnership.

Attachments

2.5 Technical questionnaire

see the following documents:

Czech version: MERLIN_-_CZ__tecnico.pdf

German version: MERLIN_-_DEU__technisch.pdf

Italian version: MERLIN_-_ITA__tecnico.pdf

2.6 Interview guidelines

Leitfaden für Interviews in Ergänzung zur Fragebogenerhebung, technischer Teil

Vor dem Interview wird der/die TeilnehmerIn gebeten schriftlich einen Kurzfragebogen mit personenbezogenen Daten (vgl. Fragebogen) sowie die Einverständniserklärung/Datenschutzbestimmung auszufüllen.

Inhaltliche Einleitung des Interviews:

[Begrüßung und Dank für die Teilnahme, plus einige Informationen zum Projekt]

... Im Rahmen des Projektes MERLIN werden schriftliche Lernerproduktionen aus standardisierten Sprachtests über eine kostenfreie Online-Plattform zur Verfügung gestellt. Diese authentischen Lernertexte wurden zuverlässig auf den GeRS bezogen und umfassen die Niveaustufen A1 bis C1. Sie werden für sprachliche Merkmale, die entsprechend des GeRS charakteristisch für Lernersprache sind, ausgezeichnet und enthalten Informationen zu dem Profil des Testnehmers (wie z.B. dessen Erstsprache). Um die Bearbeitung der Texte und die Gestaltung der Plattform möglichst gut auf die Bedürfnisse von potentiellen Nutzern, wie Ihnen, abzustimmen, möchten wir Sie in diesem Interview um einige Einblicke dazu bitten, wie Sie sich vorstellen könnten mit aufbereiteten Lernertexten zu arbeiten – aufbereitet in dem Sinne, dass die Texte bewertet, kontextualisiert und interessante Aspekte darin markiert wurden. Alles klar?

Use cases

1. In welchen Arbeitssituationen wäre es für Sie hilfreich auf Beispiele von Lernertexten für die unterschiedlichen GeRS-Niveaus zugreifen zu können?
2. Wie würden Sie diese Textbeispiele verwenden? Oder allgemeiner, wie würden Sie mit einer großen Datenbasis an Lernertexten, wie sie von MERLIN bereitgestellt wird, arbeiten wollen?

Gestaltung der Plattform - Funktion und Darstellung

3. Die MERLIN Online-Plattform wird eine Such- und Arbeitsumgebung für die Verwendung von authentischen Lernerproduktionen zu den GeRS-Niveaus anbieten. Wie würde nach Ihren Vorstellungen das Arbeiten auf einer solchen Plattform aussehen?
4. Wie würden Sie die Daten durchsuchen und analysieren wollen? Welche Wünsche haben Sie bezüglich der Anzeige der Daten? (*Filtern nach Metadaten/Merkmalen, Hervorheben/Ausblenden von Merkmalen*)
5. Sind Sie vorwiegend an vollständigen Textproduktionen interessiert, oder ist es für Ihre Arbeit eher - oder auch - relevant einzelne Merkmale von Lernersprache im Detail oder außerhalb

ihres Kontextes zu betrachten, gegebenenfalls auch im Vergleich mehrerer Texte? (*Vergleiche von Merkmalen über Textgruppen, Auflistung von Merkmalsausprägungen*)

6. Würden Sie von Hilfedateien in Form von ausführlichen Beschreibungen, Kurzanweisungen in Pop-Ups und Videos Gebrauch machen?
7. Wäre es für Sie hilfreich auf der Plattform Informationen zu den Texten speichern zu können und individuell mit den Texten auf der Plattform arbeiten zu können?

Arbeiten im Team

8. Arbeiten Sie gemeinsam mit KollegInnen in Kontexten, in denen Beispieltexte zu den GeRS-Niveaus Ihre gemeinsame Arbeit unterstützen könnten?
9. Wie stellen Sie sich die Arbeit im Team auf einer Plattform, wie MERLIN sie anbieten will, vor? Gibt es spezielle Funktionen, die Ihnen die Arbeit in der Gruppe erleichtern könnten?

+ see document: **Leitfaden-technischANDlinguistisch_v3_FINAL.pdf**

2.7 Summary report

see document: Summary_report_Anlage2.7.txt

2.8 Invitation letters and privacy documents

see documents:

final_Einleitung_Teil2-DE-IT-CZ.pdf

Privacy_CZ_Teil2.pdf

Privacy_DE_Teil2.pdf

Privacy_ITA_Teil2.pdf

Privacy_DE_Interviews_technisch.pdf