WP 4 – User modeling

Deliverable 3.1 - Report on user relevance

Part 1 – **Content aspects of the MERLIN platform**

1 Report on content aspects of the MERLIN platform

1.1 Introduction

The work carried out in WP4 user modeling aims at collecting
·   users' needs concerning the features of learner language that are considered most relevant,
·   users' needs concerning the illustration of CEFR levels ,
·   data as a basis for the development of experiential, user-based indicators describing L2-competence (WP 5),
and at guaranteeing the adequacy of the tool on a content level.

In order to achieve this goal, an online survey was organized in April 2012. The questionnaire was pre-tested between 20th and 23rd March 2012 with 3 teachers of German and 1 teacher of Italian. Few slight modifications were made.

The respondents were asked some questions about different aspects (see attachments 2.5) such as:
  –   their knowledge and working experience with the CEFR and with language competence testing (Section 2)
  –   their meaning about language learning difficulties and the criteria for evaluating learners' productions (Section 3)
They were also asked to evaluate a learner's text with the aid of the CEFR scales/descriptors and to judge the task of applying the CEFR scales/descriptors (Section 4).

Section 1 dealt with the respondents' profile.
The questionnaire was written in all three official languages of the MERLIN project (Czech, German and Italian) and it opened with a series of questions about the respondents' profiles.

A total of 408 people looked at the questionnaire, but only 115 of them completed it (Table 1).[1] Women were, not surprisingly, more numerous than men in each group (Table 2). Many respondents did not tell their gender.

	language used in the survey			
	CZ	DE	IT	Total
not completed	119	113	61	293
completed	46	31	38	115
Total	165	144	99	408

Table 1: Overview completed questionnaire per language.

	Gender			Total
	CZ	DE	IT	
woman	39	34	54	127
man	5	12	13	30
missing	121	98	32	251
Total	165	144	99	408

Table 2: Gender of respondents.

The respondents are born between 1934 and 1986 and they represent all target groups distributed as shown in Figure 1**Figure 1Figure 1**.[2] Professional profiles have been resumed in four categories, as shown in Figure 2. The distribution of working languages is shown in Figure 3.

---

[1] For the present analysis, only completed questionnaires have been taken into consideration.
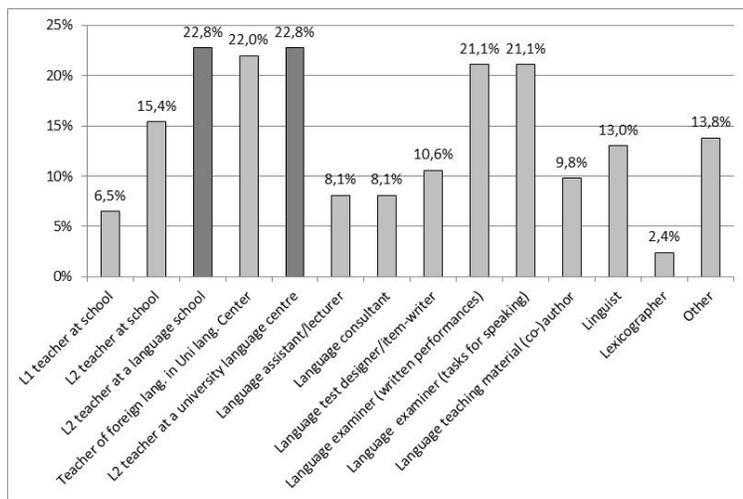
**Figure 1: profession of respondents (multiple answers allowed).**
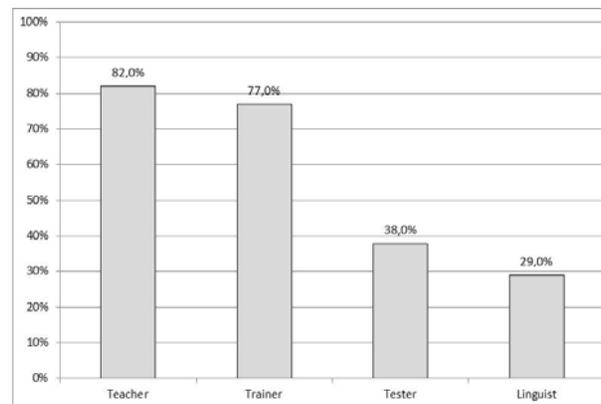
**Figure 2: categories of professional figures.**
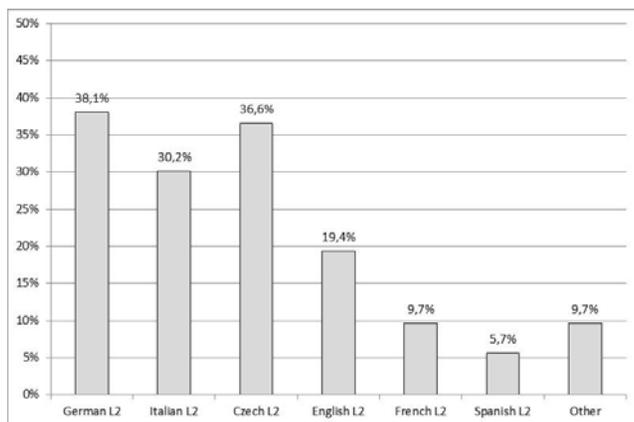
Figure 3: Working languages of respondents

	German L2	Italian L2	Czech L2	English L2	French L2	Spanish L2	Other
Czech	4,30%	0%	97,80%	15,20%	6,50%	0%	13%
German	93,50%	9,70%	0%	32,30%	12,90%	6,50%	3,20%
Italian	42%	90%	0%	18%	13%	13%	13%

Table 3: distribution of working languages/language groups.

---

[2] Respondents were allowed to select more than one profession.

1.2 Section 2: experience with the CEFR

Few respondents (9%) have had no experience with the CEFR. The others know the instrument and many of them employ it in their working routine.
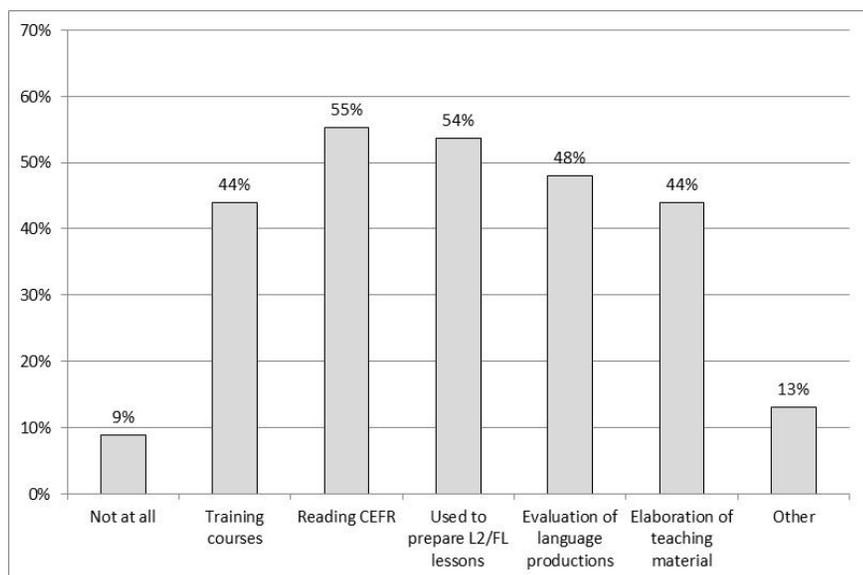

Figure 4: knowledge and working experience with the CEFR.

Overall, few Czech and German respondents are scarcely acquainted with the CEFR scales/descriptors (7% among the Czech ones and 16,2% among the German ones (Figure 5) - answers below the central value "average" -). Among the Italian respondents this percentage increases up to 26,3%, nevertheless, those who are well acquainted with the CEFR scales/descriptors represent almost the half of the sample (42,6%). The German sample holds the higher percentage of CEFR scales/descriptors well acquainted (54,9%) and the Czech the lower one (37%).
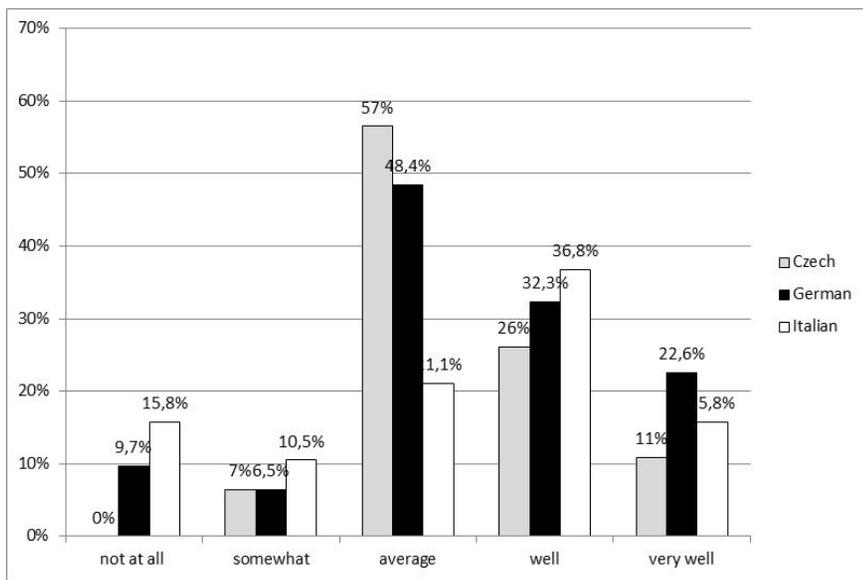
Figure 5: level of acquaintance with the CEFR scales/descriptors

76% of the Czech respondents have already worked with language tests as well as 71% of the German and 42% of the Italian respondents. The test they have worked with are those from TELC, Goethe Institut, Toefl, Cambridge, SZk, UJOP, STANAG, cles, DSH, Cils/Celi, dele, Unicert, dialang, Plida etc.
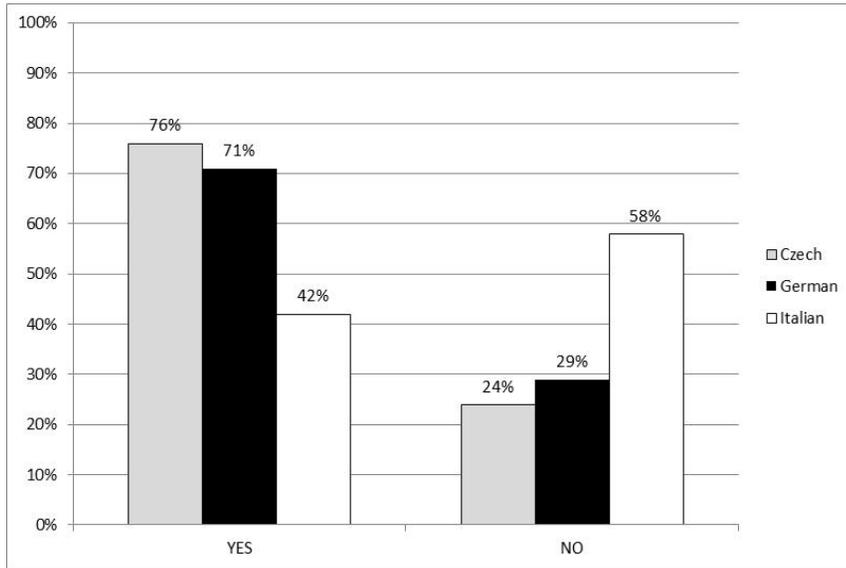
Figure 6: working experience with language tests.

2.2 Results of questionnaires[3]

Section 3: Language learning difficulties

Respondents have been asked which are the difficulties of second language learning as concerning vocabulary, grammar, coherence/cohesion, orthography and sociolinguistic appropriateness. They gave at first open and then a set of closed answers.

In the **closed answers** to the question "What do you evaluate in learners' texts?" as for **vocabulary** they signaled most frequently *vocabulary range* and *adequacy of terms*. Italian respondents also pointed out the *meaning of content words*, German respondents the *meaning of collocations* and Czech respondents the *adequacy of collocations*. (cfr. Figure 7).

NB: single words (terms) appears to be perceived as more relevant than fixed expressions/collocations.

To the question "What do you think is most difficult when learning a foreign/second language?" they spontaneously mentioned (**open answers**) *vocabulary range* (ITA: 2,6%, DE: 9,7%, CZE: 2,2%), *meaning of content words* (CZE: 8,7%), *of collocations* (DE: 9,7%), of *polysemic word/synonyms* (DE: 3,2%) and of *idiomatic expressions* (ITA: 2,6%) and false friends (ITA: 5,3%, DE: 6,5% CZE: 4,3%) (cfr. Figure 8a/b/c).

NB: false friends and interferences are perceived as important (cfr. also mentioned when working with descriptors).

[3] See attachment 2.7 for a summary report of the answers.
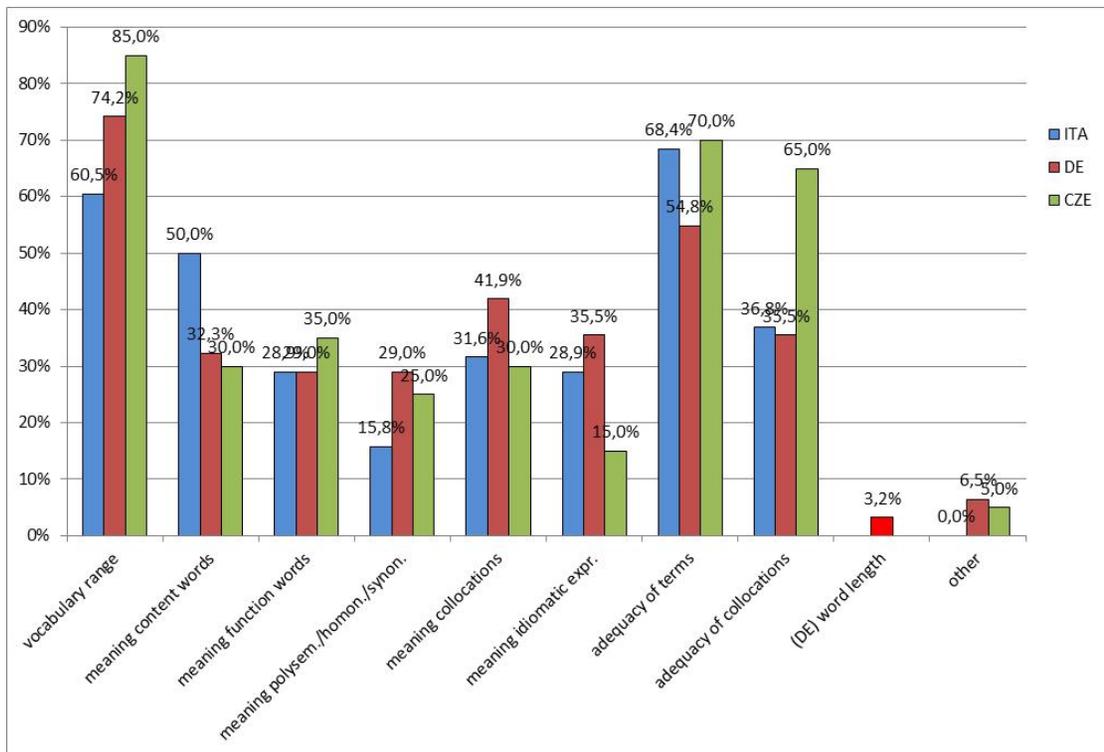
Figure 7: closed answers – aspects to evaluate in a learner's text (vocabulary).

Italian

German

Czech
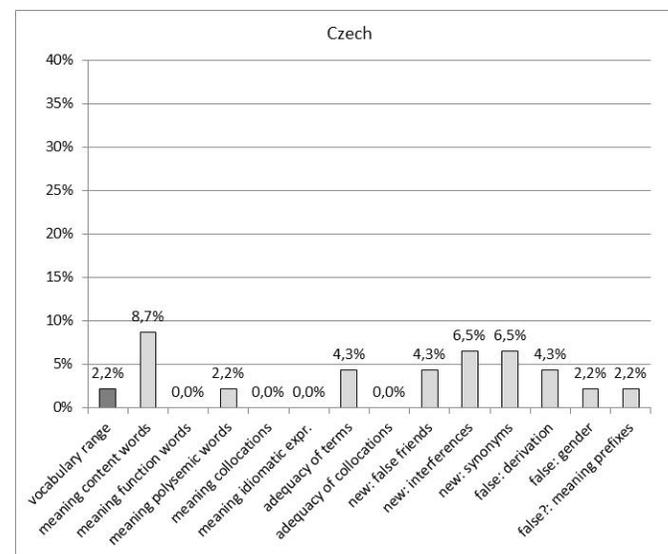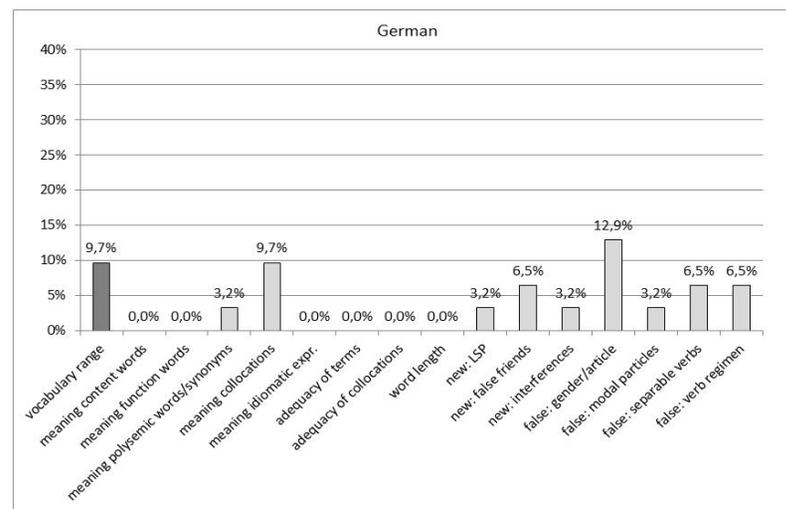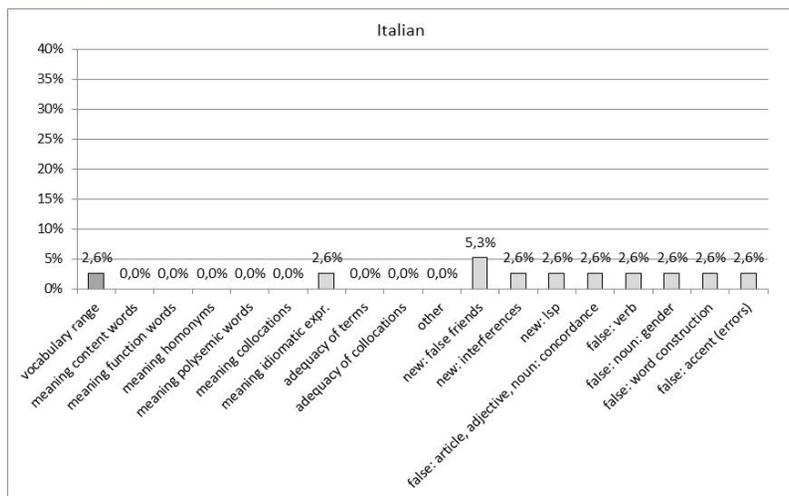
Figure 8 a/b/c: open questions – difficult aspects in language learning (vocabulary).

Grammar

As for **grammar** respondents (**closed answers**) chose, for all three languages, the category *plural formation*, while the scores for almost all other categories showed quite big differences between the three languages: i.e. Czech respondents pointed out the problematic category *verb conjugation*, while Italian respondents underlined the importance of the *accordance adjective-noun* and German that of *noun and its article* (gender).
In general, respondents focused more on morphology than on syntax. They pointed out the importance of declension/conjugation; word formation (derivation/compounding) seems to be less important; sentence length and sentence complexity are not perceived as aspects linked to each other (cfr. also report on the technical part).
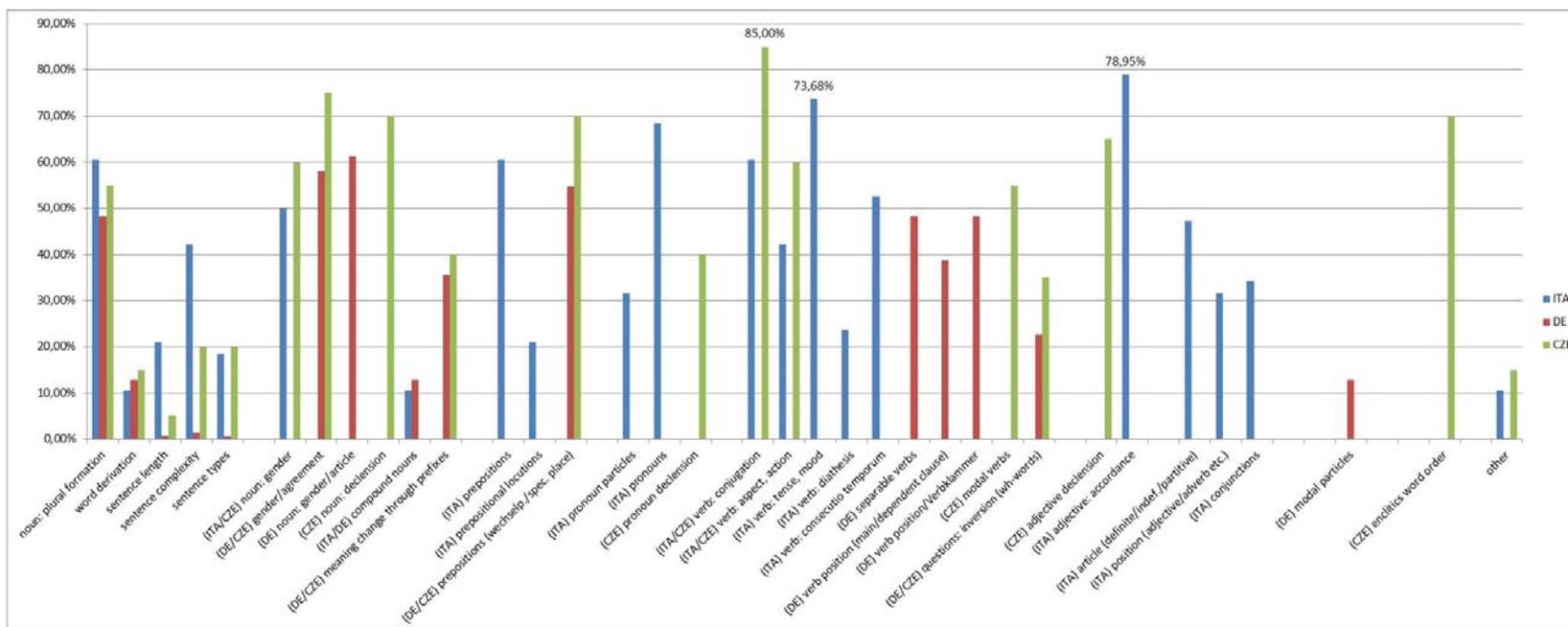

Figure 9: closed answers – aspects to evaluate in a learner's text (grammar).

In the **open answers**, respondents pointed out the importance of categories which are "traditionally" problematic when learning that language: i.e. *prepositions*, *verb tenses* and *moods* and *pronouns* for Italian, *noun/article accordance*, *adjective* and *noun declension* and *irregular verbs* for German, *declension* and *verb aspect forms* for Czech (Figure 10a/b/c).
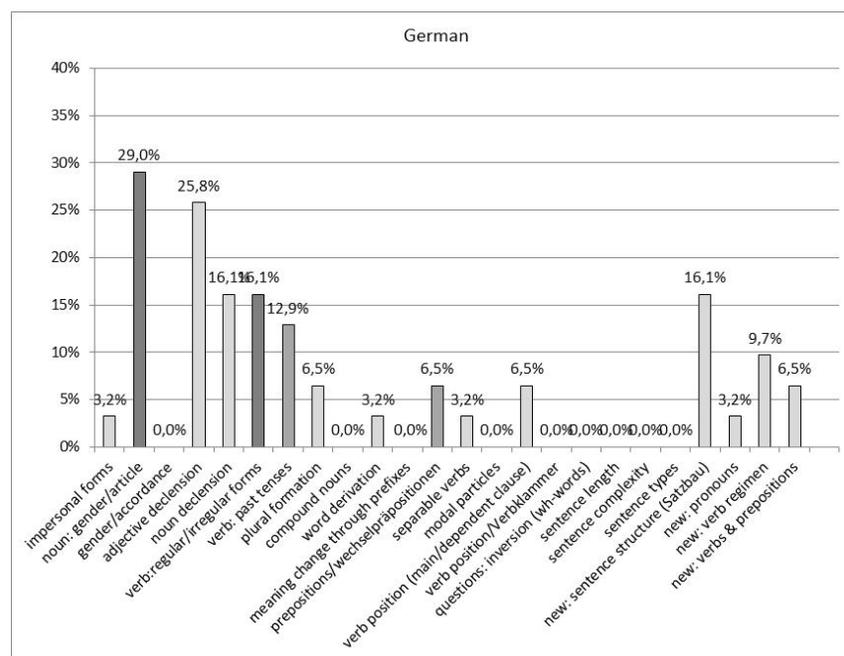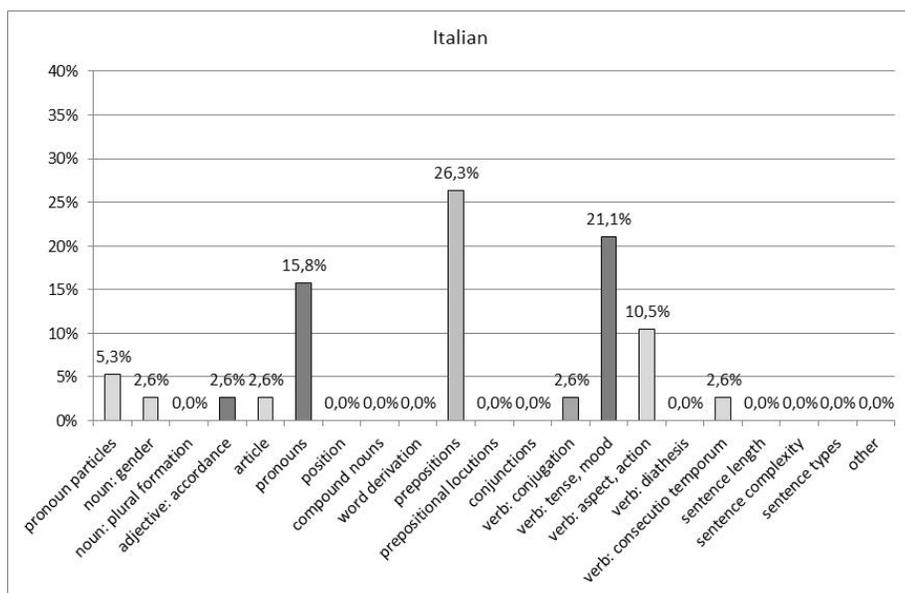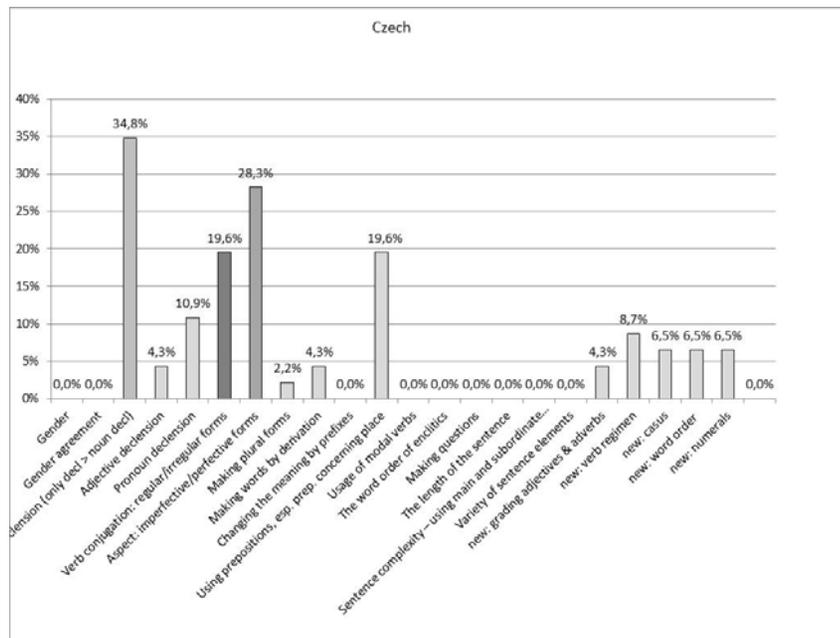
Figure 10 a/b/c: open questions – difficult aspects in language learning (grammar).

Coherence/cohesion & other text characteristics

As for __coherence/cohesion__, respondents' closed answers are quite inhomogeneous (cfr. Figure 11). Czech respondents seemed more compact in their answers: the vast majority of them pointed out the difficult aspect of bringing cohesion through connectors/verb tense or mood and the importance of the thematic structure in discourse. Italian respondents highlighted the importance of speech acts, while German respondents distributed their answers among the different categories without particular peaks.
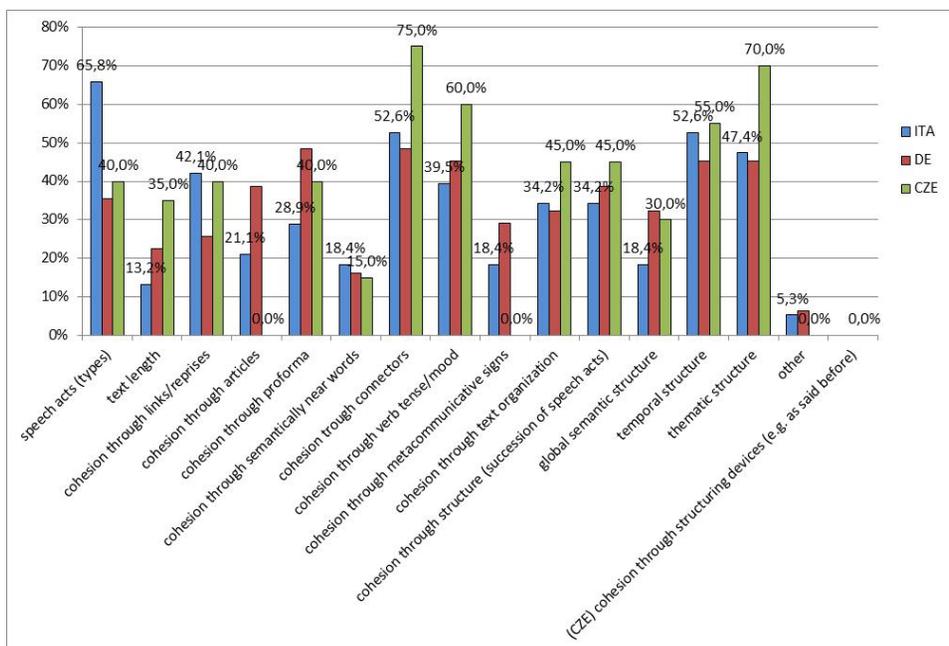
Figure 11: closed answers – aspects to evaluate in a learner's text (coherence/cohesion).

Spontaneously, respondents mentioned above all connectors (ITA: 10,5%, DE: 12,9%, CZE: 4,3%; Figure 12) as crucial difficulty in second language learning. For this category we collected very few answers, a fact that suggests that cohesion/coherence is a much less clear concept than others (see also the "false" categories named, i.e. they do not pertains to coherence/cohesion). German respondents also mentioned the deixis aspect but this mainly pertains to oral skills.
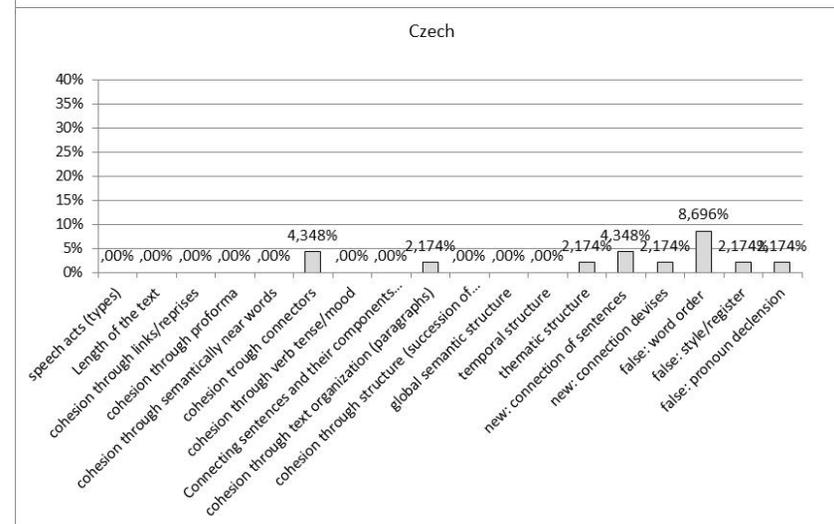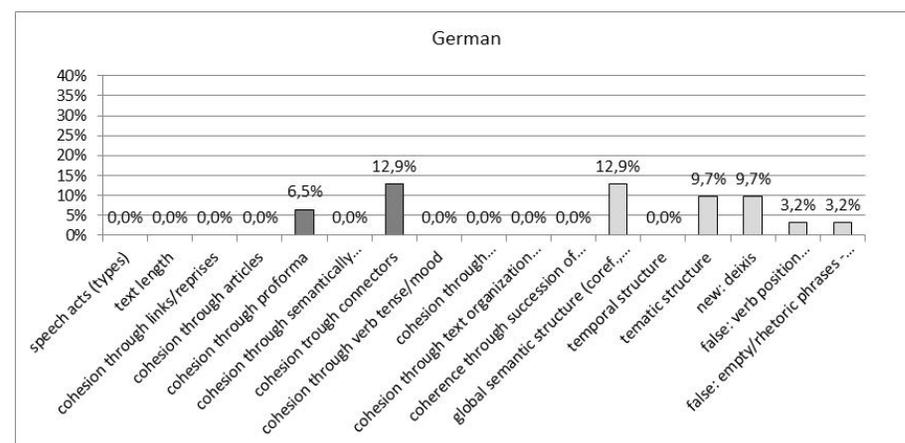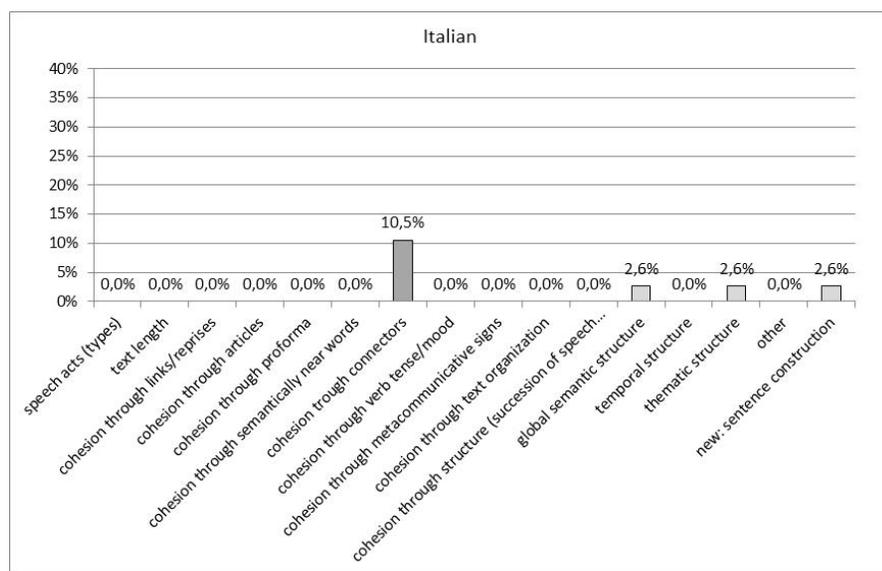
Figure 12 a/b/c: open questions – difficult aspects in language learning (coherence/cohesion).

Orthography

As for **orthography**, besides the punctuation problems, respondents obviously selected the most critical aspects related to their language: low/upper cases for German, diacritics, problematic graphemes (also related to the confusion between the Roman and the Cyrillic alphabets) for Czech and accents and apostrophes for Italian (Figure 13). The open answers confirmed the closed ones (Figure 14 a/b/c).
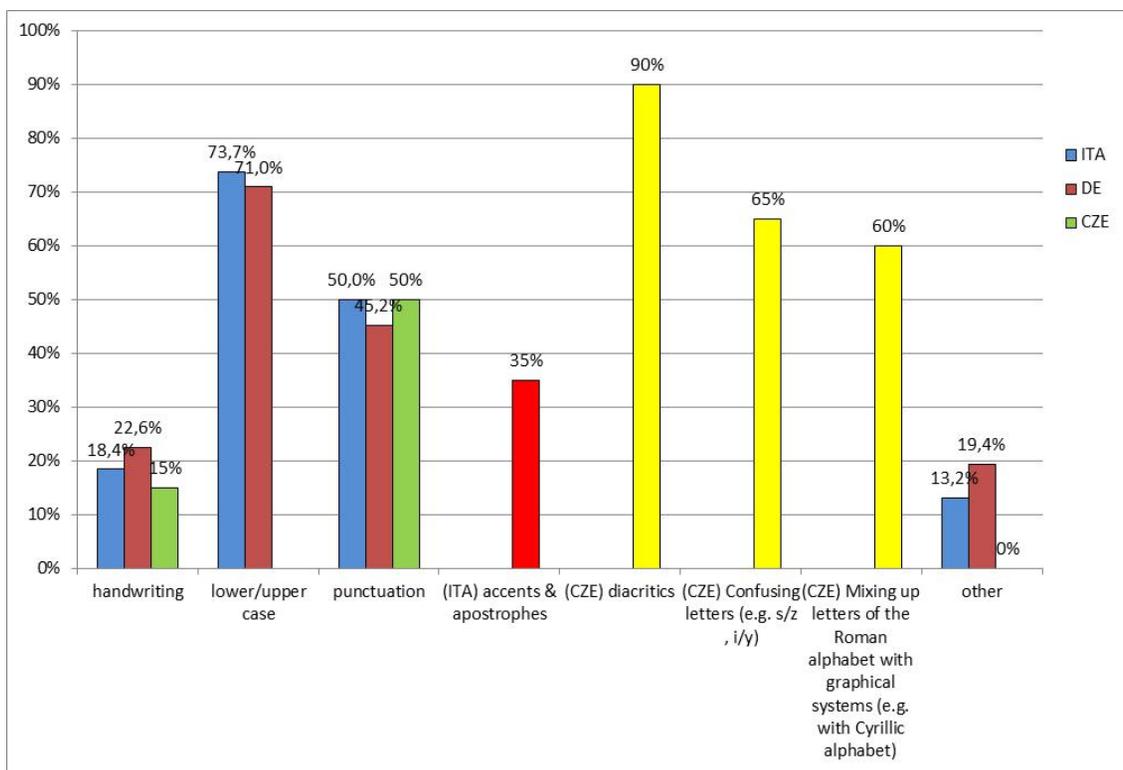
Figure 13: closed answers – aspects to evaluate in a learner's text (orthography).
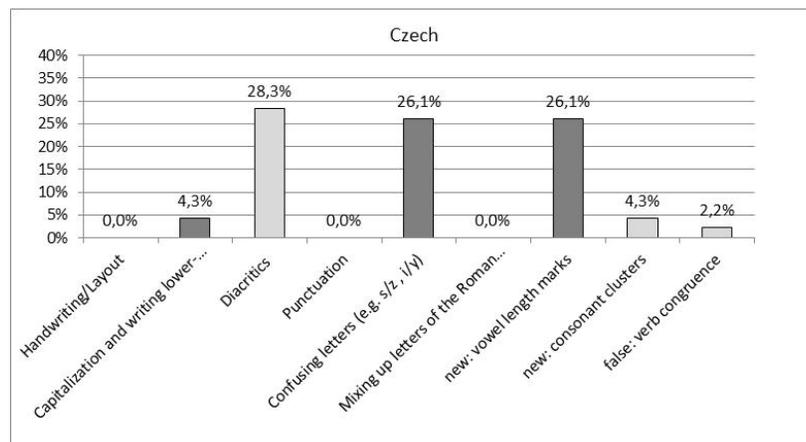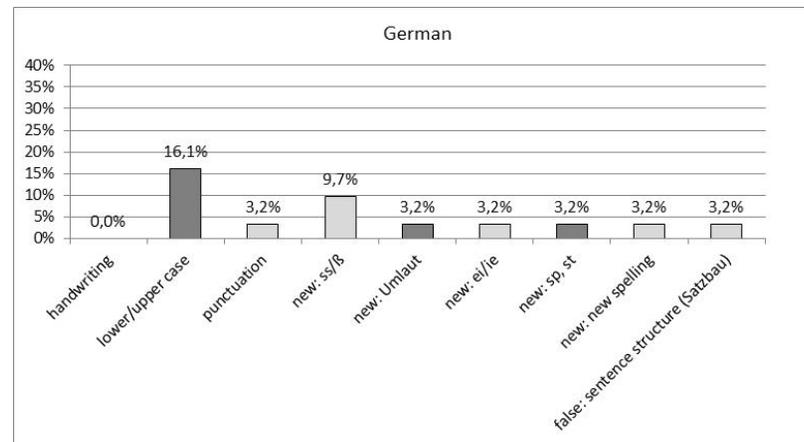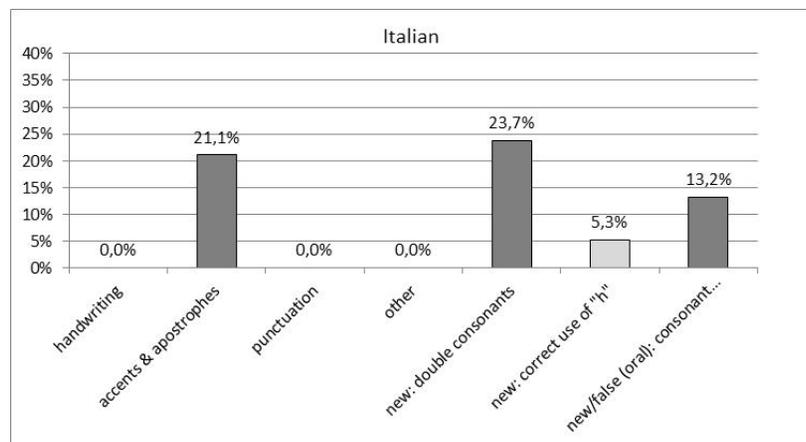
Figure 14 a/b/c: open questions – difficult aspects in language learning (orthography).

Sociolinguistic appropriateness & text genre characteristics

As for **sociolinguistic appropriateness/text genre characteristics**, Italian respondents' answers are distributed homogeneously among the given categories, while German respondents focused – both in the closed as well as in the open answers - on the *text/genre appropriateness* and Czech respondents on the *opening/closing formulas* and on the *style and register* adopted in writing (Figure 15). In the open answers the respondents of all three languages also mentioned the importance of *politeness convention* (Figure 16 a/b/c).
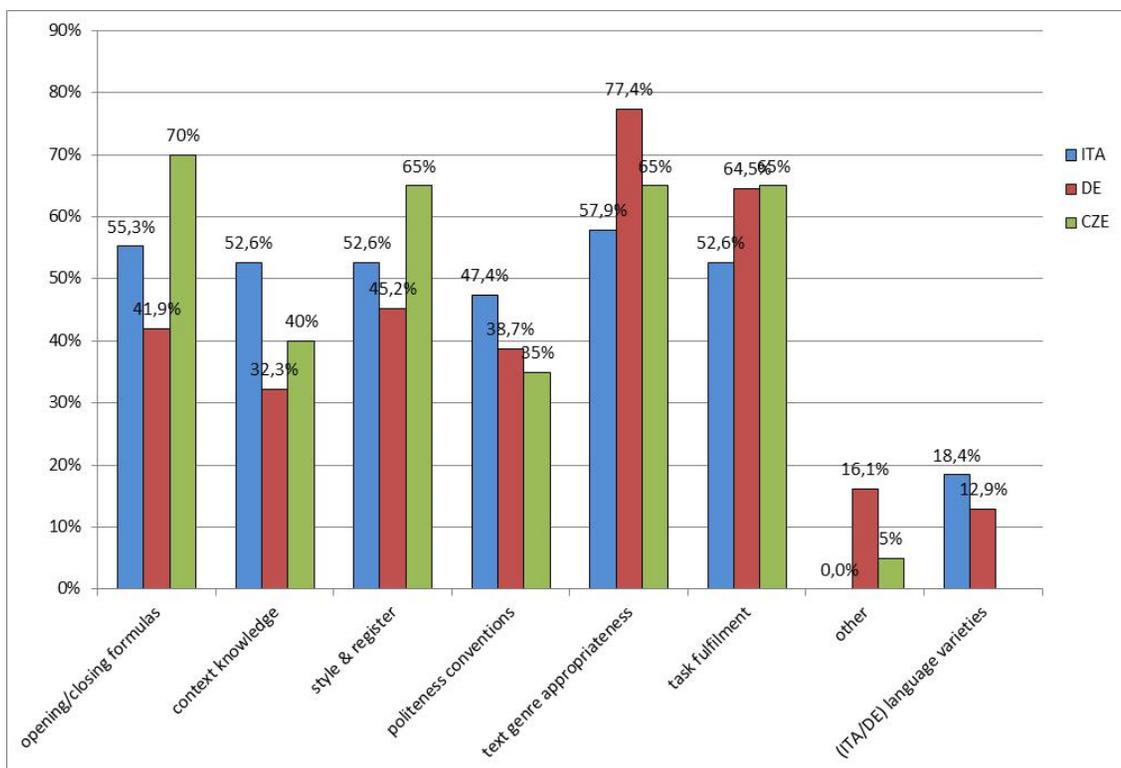
**Figure 15: closed answers – aspects to evaluate in a learner's text (sociolinguistic appropriateness).**
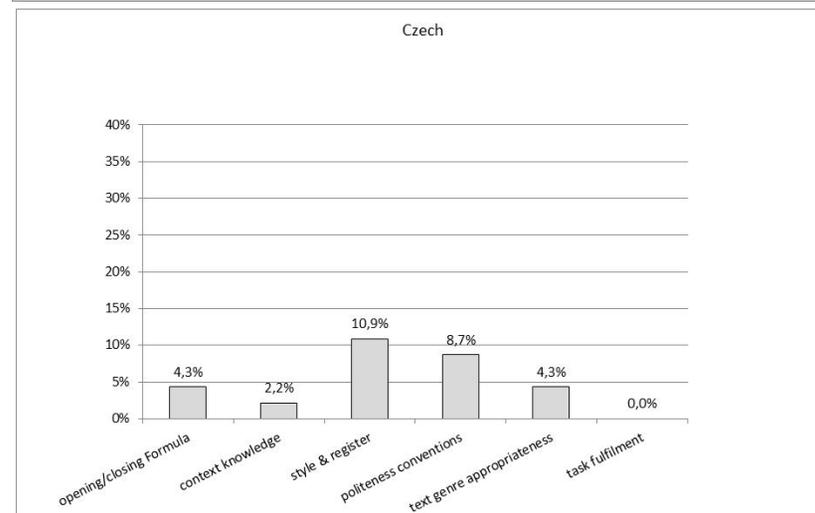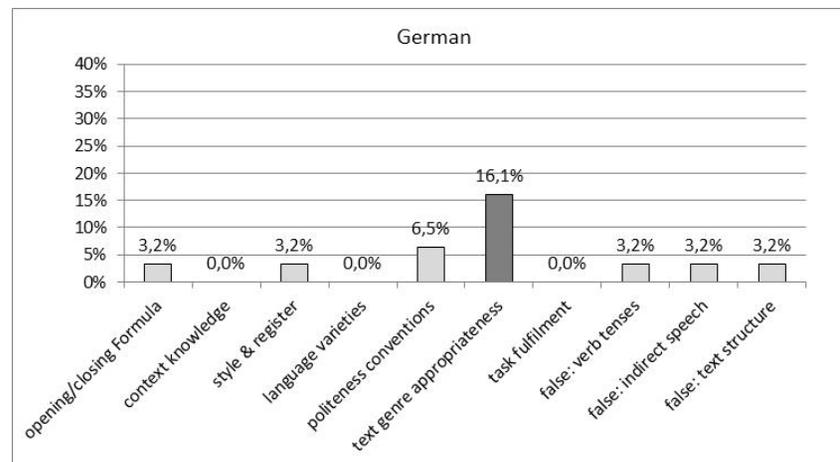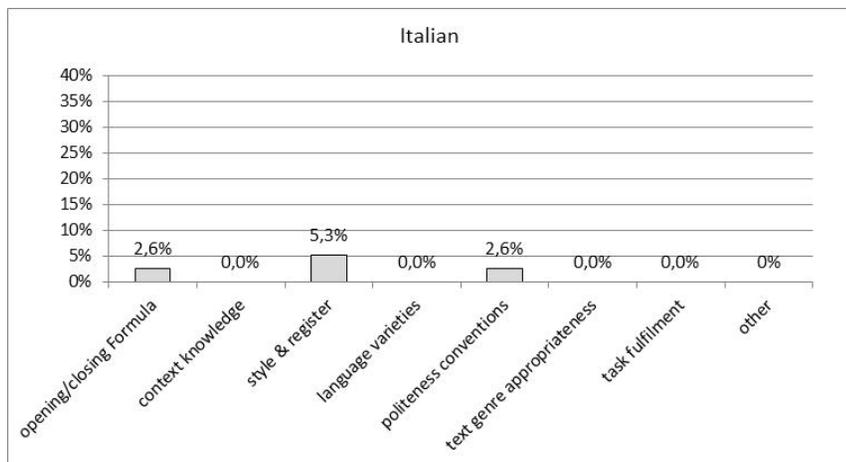
Figure 16 a/b/c: open questions – difficult aspects in language learning (sociolinguistic appropriateness).

Section 4: Evaluation of a learner's text and impressions about working with the CEFR scales/descriptors

Respondents have been asked to evaluate a learner's text (which had previously been evaluated by members of the MERLIN team within other projects). They were automatically assigned one of two texts and three of six CEFR scales (vocabulary range, vocabulary control and sociolinguistic appropriateness or grammatical accuracy, orthographic control and cohesion/coherence). After the task was completed they set forth the critical but also the useful aspects of the CEFR scales and they explained how they had met their decision about the text.

Level of difficulty of working with the CEFR scales

In general: the "neither easy nor difficult" answer tends to be, not surprisingly, the most chosen by respondents.
The scales that seem to be a bit more problematic are the sociolinguistic appropriateness (difficult = 27,1%, easy = 27,1%, Figure 22 a/b), the orthographic (difficult = 21,4%, cfr. Figure 20 a/b) and the vocabulary control one (difficult = 20,3%, though the percentage of respondents who consider easy working with it is quite high (40,7%), cfr. Figure 18 a/b).
The scale which appear to be the most easy to work with is the vocabulary range one (easy = 40,7%, difficult = 10,2%, cfr. Figure 17 a/b)
The most "neutral" scales are the coherence/cohesion (neither easy/nor difficult = 69,6%, cfr. Figure 21 a/b) and the grammatical accuracy one (neither easy/nor difficult =66,1%, Figure 19 a/b). For a summary look at Figure 23.
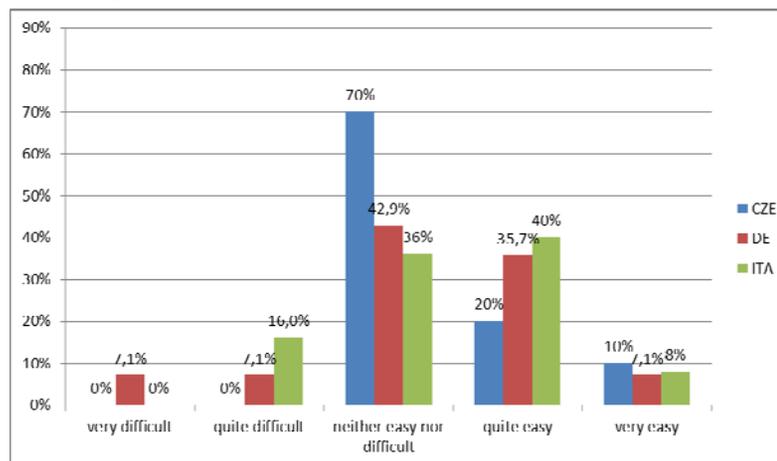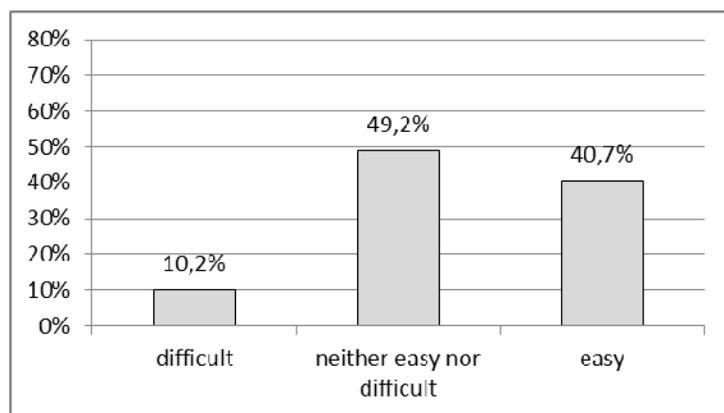
Figure 17 a/b: level of difficulty of working with the Vocabulary range scale.
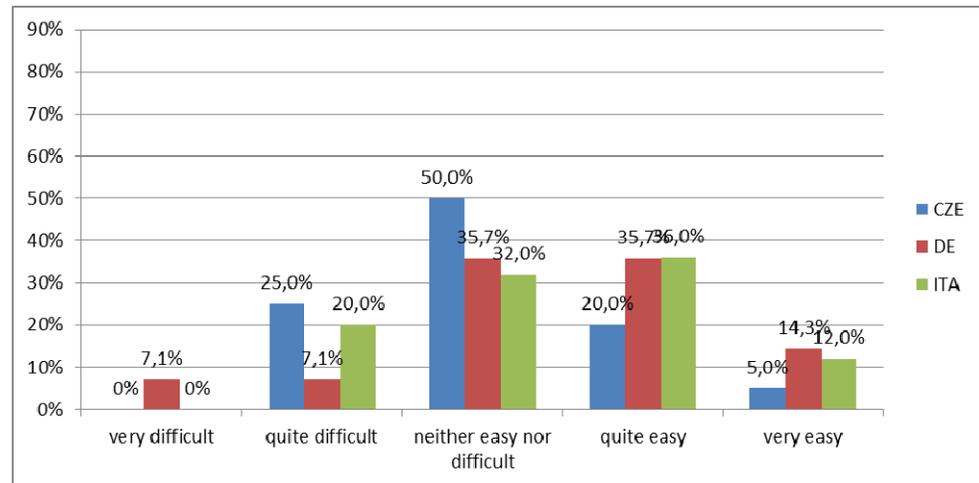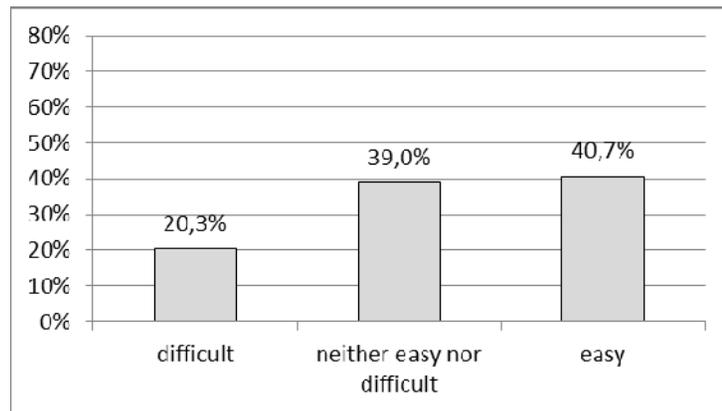
Figure 18 a/b: level of difficulty of working with the vocabulary control scale.
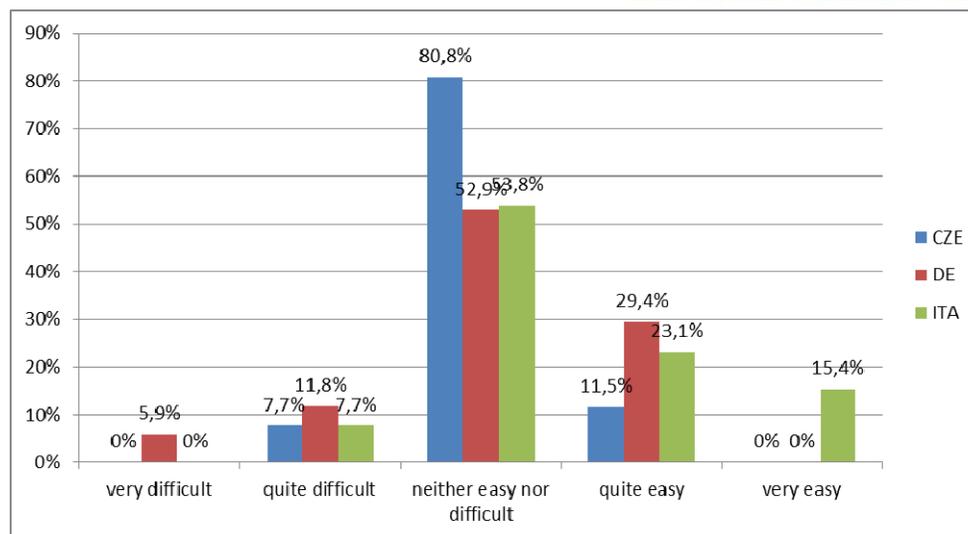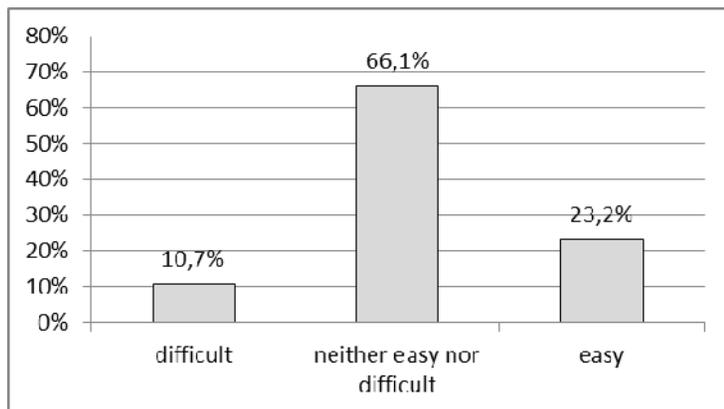
Figure 19 a/b: level of difficulty of working with the grammatical accuracy scale.
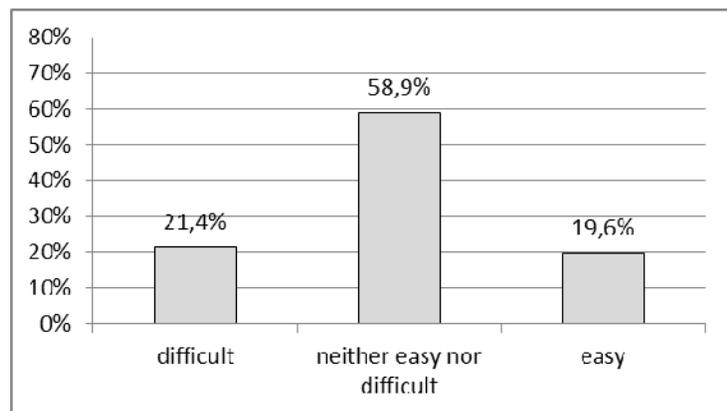
Figure 20 a/b: level of difficulty of working with the orthographic control scale.
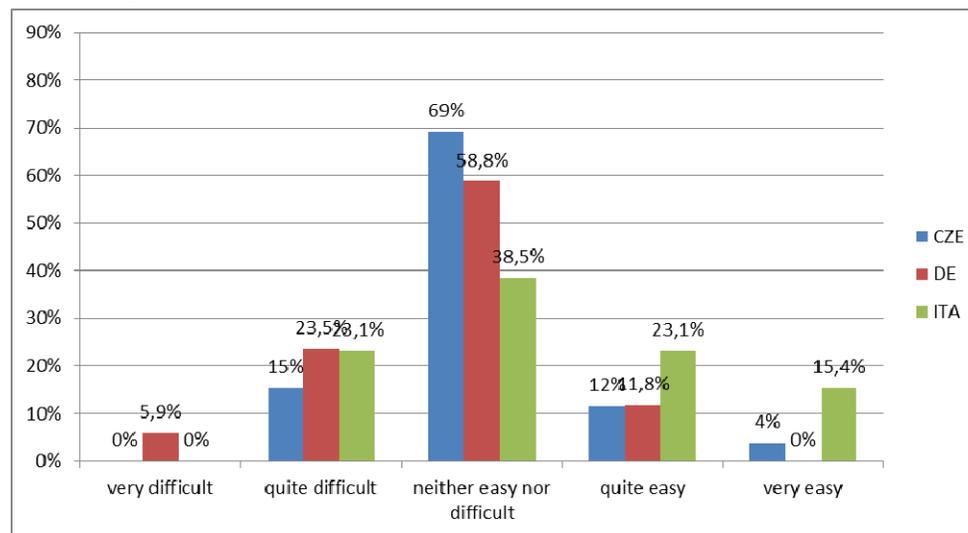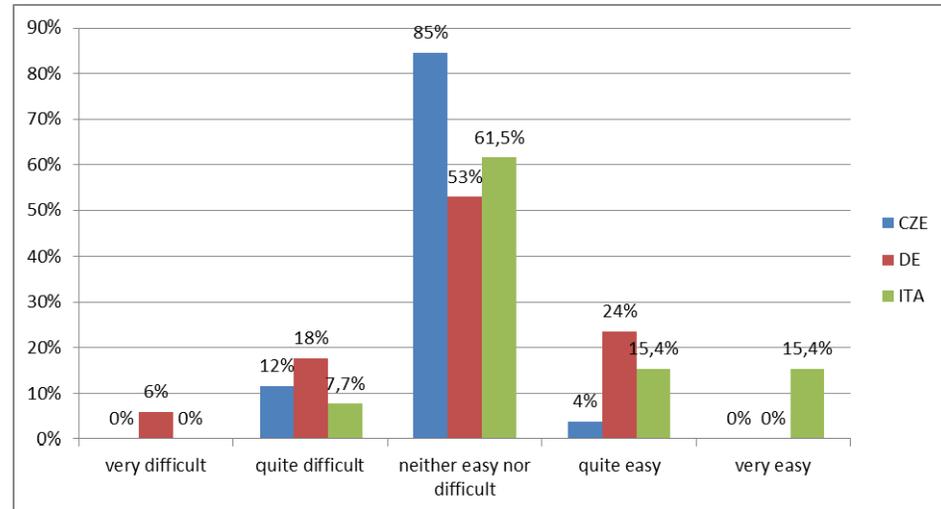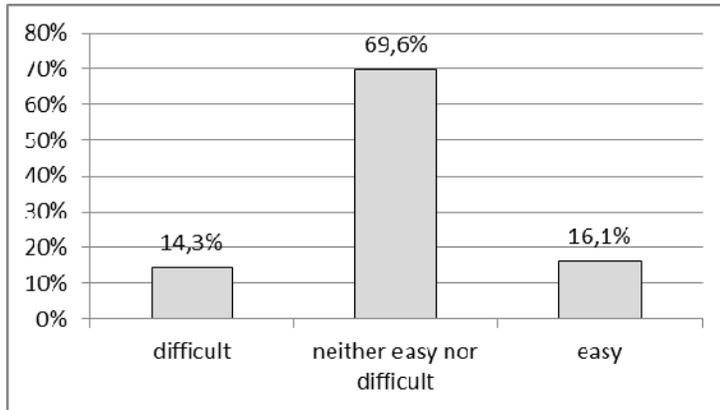
Figure 21 a/b: level of difficulty of working with the Cohesion/coherence scale.
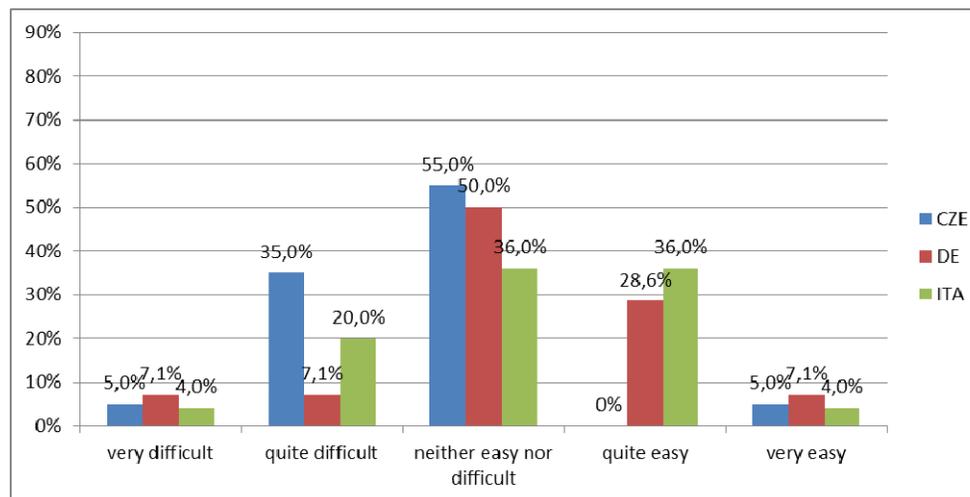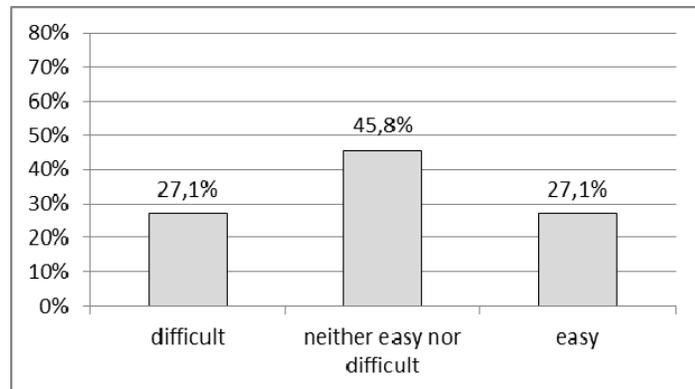
Figure 22 a/b: level of difficulty of working with the sociolinguistic appropriateness scale.

Working with CEFR	Coherence	Orthography	Grammar	Voc. range	Voc. control	Sociling.
Very easy	15%	15%	15%	8%	12%	4%
Likely easy	15%	23%	23%	40%	36%	36%
Adequate	62%	39%	54%	36%	32%	36%
Likely hard	8%	23%	8%	16%	20%	20%
Very hard	0%	0%	0%	0%	0%	4%

Figure 23: subjective evaluation about working with the CEFR scales (all respondents).

Critical and useful aspects of the CEFR descriptors/scales

As for the critical and useful aspects of the CEFR descriptors/scales, respondents selected and listed a series of elements they isolated in each scale descriptor as can be seen below. Furthermore, they expressed their doubts, problems and desiderata about the scales themselves.

As for the vocabulary range/control scales (Table 4 a/b and Table 5 a/b), there is an incongruence: the formulations which are said to be useful in the vocabulary range scale are at the same time those who perplexed/puzzled the respondents (cfr. „sufficient", „basic", „broad").
The hint at idiomatic expressions is said to be useful, however respondents do not pay great attention to this aspect when they evaluate learners' texts (rather adequacy of terms, cfr. Figure 8).

Level	Helpful formulas from levels	DE	IT	CZ	TOTAL
C2	very broad vocabulary	1			1
	idiomatic expressions	1		1	2
C1	broad vocabulary	1			1
	avoidance strategies seldom		1		1
B2	broad vocabulary	1			1
	own field			1	1
	most general topics			1	1
	avoidance of repetitions			1	1
	lexical gaps	1			1
	circumlocution			1	1
B1	sufficient vocabulary	2			2
	everyday life	1			1
	some circumlocution	1			1
A2	sufficient vocabulary	1	1		2
	expression of basic communicative needs	2		1	3
A1	isolated words			1	1
	isolated phrases			1	1

Categories	Formulations/aspects/opinions	DE	IT	CZ	TOT
criticalities	C1 description		1	1	2
	difficult to identify circumlocution or hesitation	1		1	2
desiderata	need to expand the formulations			1	1
	more detailed/concrete descriptions (concrete examples only B1)	1		1	2
understanding	lexical limitations cause repetition and even difficulty with formulations			2	2
	adjectives (broad lexical repertoire; sufficient) & formulas (elementar comm. Needs)	3		1	4
	good range of vocabulary vs. sufficient vocabulary; express him/herself with some circumlocutions vs. Can vary formulation to avoid frequent repetition, but lexical gaps		1	1	2

Table 4 a/b: helpful vs. critical elements of the vocabulary range scale.

Level	Formulations/aspects/opinions	DE	IT	CZ	TOTAL
C2	consistently correct			1	1
	consistently appropriate use of vocabulary		1	1	2
	minor slips				0
	no significant vocabulary errors			1	1
C1	generally high lexical accuracy		1	1	2
	some confusion/incorrect word choice		1	1	2
	don't hinder communication	1	3		4
B1	good control of elementary vocabulary	1	2		3
	major errors > complex thoughts	1		4	5
A2	narrow repertoire	2	1		3
	everyday needs	2			2

Categories	Formulations/aspects/opinions	DE	IT	CZ	TOTAL
criticalities	A1 no descriptor available		1	1	2
	not clear difference btw voc. range and voc. control scales			1	1
	concepts are not unambiguous	1			1
	use of "vocab use" vs. "vocab control"	1			1
	levels describe errors rather than quality of text production		1		1
understanding	minor slip; generally high; etc.		2	1	3
	more complex thoughts		1		1
	difference btw. elementary vocab/narrow repertoire (B1/A2)	1			1

Table 5 a/b: helpful vs. critical elements of the vocabulary control scale.

AS for the grammatical accuracy scale, the attention paid by the respondents to „error measuring" (cfr. the usefulness of the formulations which „count" the errors: rare, difficult to spot, systematic basic; Table 6 a/b) can be probably traced back to the traditional assessment practice at school which sometimes represents an obstacle in applying the can-do approach of the CEFR.

Categories	Formulations/aspects/opinions	DE	IT	CZ	TOTAL
C2					0
C1	errors are rare			1	1
	errors are difficult to spot			1	1
B2	relatively high degree of grammatical control			1	1
	errors don't lead to misunderstanding		2	3	5
B1	reasonable accuracy		1		1
	repertoire of frequently used "routines"		1	1	2
	repertoire of frequently used patterns			1	1
	link to predictable situations		1	2	3
A2	some simple structures correctly used	1			1
	systematic basic mistakes	1		1	2
	mix up tenses, forget to mark agreement			2	2
	usually clear what one is trying to say	1	1		2
A1					0
					0

Categories	Formulations/aspects/opinions	DE	IT	CZ	TOTAL
criticalities	lack of specification (types of mistakes specified/listed only at A2 level)	1		1	2
	incoherence: seriousness of mistakes (not mentioned at B1), frequency of mistakes (not mentioned at B1/2)			1	1
	correctness/comprehensibility not always both taken into consideration (incongruent)	1	1		2
	tendence to negative formulations use (?)	1			1
understanding	(B1) uses reasonably accurately a repertoire of frequently used routines and patterns associated...			1	1
	mistakes lead to misunderstandings			1	1
	consistently, seldom, relatively...; also "mistake"; "learnt repertoire" (A1); (C2) "attention mybe engaged otherwise"	2	1	1	4

Table 6 a/b: helpful vs. critical elements of the grammatical accuracy scale.

About the cohesion/coherence scale, the few respondents who stated an opinion pointed out the need for further specifications.

Level	Helpful formulas from levels	DE	IT	CZ	TOTAL
C2	variety			1	1
	wide range			2	2
C1	controlled use - organis. patterns			2	2
B2	limited number			2	2
	jumpiness in long contribution	1		1	2
B1	connected sequence of points	1		1	2
	linear sequence of points	1	1		2
	series of shorter simple elements		1	1	2
	link dicrete elements			1	1
A2	link group of words			1	1
	simple connectors			1	1
	(specific) connectors			1	1
A1	link words/group of words			1	1

Categories	Formulations/aspects/opinions	DE	IT	CZ	TOTAL
criticalities	overlapping [***not true??***]: B1/B2 (linear sequence of points/jumpiness in long contribution)			1	1
	C1: refers to oral texts	2			2
desiderata	further specifications needed			1	1
understanding	esp. word choice in B2 ("B2 ist eine Zumutung")	1			1

Table 7 a/b: helpful vs. critical elements of the cohesion/coherence scale.

There is an extreme uncertainty to what pertains to orthographic control among the respondents (cfr. comments : *paragraphing conventions* pertain to cohesion/coherence scale/pragmatics; it's not clear what falls under orthography).

Level	Helpful formulas from levels	DE	IT	CZ	TOTAL
C2	free of errors	2		1	3
C1					0
					0
C1 ext.	helpful (it may refers to layout/punctuation helpful)			1	1
B2	clearly intelligible			1	1
B1	generally intelligible			2	2
	continuous writing			1	1
A2	short sentences on everyday subjects			1	1
A1	copy familiar words			1	1

Categories	Formulations/aspects/opinions	DE	IT	CZ	TOTAL
	paragraphing conventions pertain to cohesion/coherence			1	1
	C2: not clear what falls under ortography			1	1
	A2/B1: different aspects assessed			1	1
criticalities	lack of more concrete descriptions (type/quantity of errors)	1	1		2
	B2/C1 refers to pragmatics	1			1
	paragraphing conventions do not pertain to the text		1		1
desiderata	further specifications			1	1
understanding	"can copy"			1	1

Table 8 a/b: helpful vs. critical elements of the Orthographic control scale.

As for the sociolinguistic appropriateness scale, respondents pointed out that no hint at text genres are given (a part from the genre "letter" but only to express its – debatable – monologue features), but they didn't mention task fulfillment and style/register which were, on the contrary, mentioned as important in the evaluating task.

Level	Helpful formulas from levels	DE	IT	CZ	TOTAL
C2	idiomatic expressions			1	1
	colloquialisms			1	1
	awareness of connotative levels of meaning			1	1
C1	wide range of idiomatic expressions			1	1
	wide range of colloquialisms			1	1
	recognition of register shifts		1	1	2
B2	without unintentionally irritating them	1			1
	without requiring them to behave other than with L1 speakers	1			1
B1	wide range of language functions	2			2
	use of most common exponents	1			1
	neutral register			1	1
	salient politeness conventions	2			2
	aware of most significant differences btw customs, usages etc.	1			1
A2	social exchanges	2			2
	everyday polite forms of greetings and address			1	1
A1	greetings and farewells, introductions etc			1	1

Categories	Formulations/aspects/opinions	DE	IT	CZ	TOTAL
criticalities	B2: (letter =) monologue not described	1		1	2
	B2 descriptor cannot be used/applied here (mainly oral registers)		2	1	3
	different comm situations/levels		1		1
understanding	"with some effort"; "aware of the most significant differences between customs, usages etc"			1	1
	B1: wide range of language functions; B2: "without unintentionally irritating them/amusing them"			1	1

Table 9 a/b: helpful vs. critical elements of the sociolinguistic appropriateness scale.

To sum up

Some general considerations:
Claims/wishes:
- more concreteness
- examples/list of phenomena

Criticism:
- vagueness of descriptors
- focus too much on oral skills
- different categories on different CEFR levels make comparisons difficult
- inconsistent terminology

Section 5: The interviews

4 expert interviews (DE 1 + 2, CZ 1 + 2)

Particular aims
- deeper understanding of users' needs and expectations
- exemplary use cases for the work with CEFR related to learner's performances
- features of learner language relevant for the own work
- experiences with CEFR related resources

Usage scenarios:
Assessment of tests:
- comparison with standardized samples (B2, C1)
- assessment of borderline performances (B1, B2, C1)
language teaching:
- comparison of own performances with assessed samples, for advanced groups (B2, C1, C2)
- (agreement between all interviews, content & technical part)

Scenario 1: aims at helping learners to set and achieve learning targets
- Learner will complete one of the tasks that MERLIN texts are based on.
- Result is compared with (non-annotated) learner texts of MERLIN (comparison and placement of learners' text)
- Looking at lists of features of learner language at level of the learner.
- Definition of learning targets.

Scenario 2: aims at defining common assessment criteria
- A MERLIN learner text without annotations is handed over to teachers/testers for rating
- Rating results are discussed in the group and compared with MERLIN rating
- Feature lists for language proficiency levels are discussed and related to different teaching/testing realities
- (= taken from interviews, technical part)

Use cases

Expectations for the use of a platform with learner performances:
- selection of samples according to CEFR levels (B2, C2)
- tracing of mistakes e.g. according to different L1 (B2, C2) = relevance of metadata
- grouping according to text genres (B2) = relevance of metadata
- working with whole texts as well as with parts of texts (B1)
- profiling of learner performances: support for the assignment of single rating criteria of one text to different levels (e.g. a B1-performance, but vocabulary range on a higher level) (B1)
- automatic analysis of learner performances (B1)
- searching for different, more or less complex linguistic means (related to different CEFR levels) that allows to do particular speech acts (B1) = functional approach
- identification of certain linguistic functional means on different CEFR levels, e.g. for building cohesion (C2, B1) = functional approach

Expectations for the use of a platform with learner performances:

Different degrees of additional information and interaction:
- plain learner texts/texts without any additional information, annotations etc.
- learner texts & assigned level & annotations & MERLIN rating criteria
- learner texts & corrections and correction comments/decision criteria
- interaction with the MERLIN team/discussion of rating decisions
- (merged from: interviews, content & technical part)

Features of learner language

- emphasis on single criteria according to the level (C1, C2, B1)
- to some extent uncertainty regarding the necessary degree of correctness vs. comprehensibility (e.g. B1)
- ranking and relevance of linguistic categories such as vocabulary, grammar etc. depends on the learner level (B1, B2, C1, C2)
- overall tendency concerning ranking:
- vocabulary at the top
- coherence/cohesion, grammar and sociolinguistic appropriateness in the midfield
- orthography at the bottom
- the terms/categories "socioling. appropr." and "coherence/cohesion" comparatively less used within the interviews/less present (B1, B2, C1, C2)

Experiences with CEFR related resources

C1 & C2 no experiences with resources such as "Profile Deutsch" (PD), B1 & B2 do (PD)

Positive:
- transfer from can-do-descriptors formulated in a general manner to concrete cases/examples (B1, B2)
- functional approach (cf. above)

Negative:
- to get lost, disoriented → suggestions: better layout, use of different colors, character fonts, graphic elements = need for a clearly structured resource (B1)
- a lot of similar formulations = need for a resource possibly without redundancies (B1)
- „Auch im Referenzrahmen, im Referenzrahmen ist es manchmal auch desorientierend, denn man kommt in so eine Art Referenzrahmentrance ... Ich kann ich kann ich kann ich kann ... Also ich kann nicht mehr (lachen)" (B1)

Attachments

2.5 Content part questionnaire
MERLIN_-_Dotazník_(A-C).pdf
MERLIN_-_Dotazník_(D-F).pdf
MERLIN_-_Fragebogen_(A-C).pdf
MERLIN_-_Fragebogen_(D-F).pdf
MERLIN_-_questionario_(A-C).pdf
MERLIN_-_questionario_(D-F).pdf

Learner texts - questionnaire:
ITA_Merlin.pdf
CZE_Merlin.pdf
ITA_Merlin.pdf

2.6 Answers – summary report
Summary_report_Anlage_2.6

2.7 Data analysis
Questionnaire_supporting_data_analysis

2.8 Interview guidelines
Leitfaden-technischANDlinguistisch_v3_FINAL.pdf

2.9 Invitation letters and privacy documents
Fragebogen_MERLIN_Einladung_with_CZ_FINAL.pdf
Privacy_DE_Interviews_linguistisch.pdf
Privacy_CZE.pdf
Privacy_DE.pdf
Privacy_ITA.pdf