MERLIN

Multilingual Platform for the European Reference Levels: interlanguage exploration in context

User Manual

Version 1
11/2014

Contents

Overview tables/graphics

PART I: Background information

PART II: User guide

References

Overview tables/graphics

Part I: Background information

1   The MERLIN project: aims and motivation

Introduction

The Common European Framework of Reference for Languages (CEFR) is the leading instrument for language teaching and certification in Europe. At its heart, although this is by no means its only contribution to standardization and vast improvements in language teaching and testing, there is the well-known system of CEFR levels, illustrated by exemplary scales. In spite of the very widespread use that is being made of these CEFR scales – there will hardly be a language test, a school curriculum, or a textbook without a reference to the scale levels – the scales are often insufficiently illustrated in terms of authentic learner data. Such concern grows even stronger when considering languages other than English (cf. e.g. Fulcher 2004, Hulstijn 2007, North 2000, Wisniewski 2014).

The project MERLIN: "Multilingual Platform for the European Reference Levels: Interlanguage Exploration in Context" aims at improving this situation by offering a contribution to the illustration and the validation of the CEFR level system. MERLIN (2012-2014) was co-financed by the European Union (Lifelong Learning Programme, 518989-LLP-1-2011-1-DE-KA2-KA2MP). MERLIN aims at researching and enhancing the empirical foundations of the CEFR scales by constructing a written learner corpus for Czech, German and Italian as L2 (cf. Wisniewski et al. 2013, Abel et al. 2014).

Background: CEFR scales

The CEFR claims to be applicable across European languages. Thus, the level descriptions had to be general, like in the example below:

B1	*Has a sufficient range of language to describe unpredictable situations, explain the main points in an idea or problem with reasonable precision and express thoughts on abstract or cultural topics such as music and films.*
	Has enough language to get by, with sufficient vocabulary to express him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work, travel, and current events, but lexical limitations cause repetition and even difficulty with formulation at times.
A2	*Has a repertoire of basic language which enables him/her to deal with everyday situations with predictable content, though he/she will generally have to compromise the message and search for words.*
	Can produce brief everyday expressions in order to satisfy simple needs of a concrete type: personal details, daily routines, wants and needs, requests for information. *Can use basic sentence patterns and communicate with memorised phrases, groups of a few words and formulae about themselves and other people, what they do, places, possessions etc.* *Has a limited repertoire of short memorised phrases covering predictable survival situations; frequent breakdowns and misunderstandings occur in non-routine situations.*
A1	*Has a very basic range of simple expressions about personal details and needs of a concrete type.*

Table 1: Exemplary CEFR scale, "General linguistic range" (CoE 2001: 110)

For MERLIN, chapter 5 scales ('communicative language competence') were used (general linguistic range | grammatical accuracy | vocabulary range | vocabulary control| orthography | coherence & cohesion | sociolinguistic appropriateness). The CEFR is downloadable from the Council of Europe website.[1]

Illustration of CEFR levels

However, it was recognized that additional, language-specific illustrations of the descriptors would be needed. In view of this demand to complement the CEFR, since 2001, the Council of Europe itself has encouraged the development of supplementary tools which better exemplify the features of single languages. One step in this direction was to instigate the publication of the *Reference Level Descriptions* (RLDs) for national and regional languages.

The tendency is that more and more RLDs tend to be based upon learner corpora, such as the English (www.englishprofile.org), but also the Italian (Spinelli/Parizzi 2010) and the Norwegian Profiles (Carlsen 2013).

While MERLIN similarly aims at illustrating CEFR levels for given languages, it differs by following, for the first time, a multilingual approach. Thus, it addresses three languages from different families (Slavic, Germanic and Romance) and supports cross-language comparisons. In addition, it is distinct from related initiatives by providing free access to the full texts, test tasks, and a wide range of linguistic and error annotations on a didactically motivated online platform. MERLIN also stands to contribute to the validation of CEFR scales.

Validation of CEFR level descriptions

The Council of Europe effort of scaling the CEFR descriptors (CoE 2001; North 2000; Schneider/North 2000) has led to immense improvements in standardization and transparency in language learning, teaching, and testing. Important decisions about language learners' lives are taken with reference to the CEFR levels. One aspect that is yet insufficiently understood is the empirical validity of the CEFR scales (Fulcher 2004; Hulstijn 2007): If scales are used to describe or rate learner language, they must reflect what learners actually do (Alderson 1991).

As CEFR levels are increasingly used in high-stakes contexts, where important decisions about people's lives depend on the interpretation of the CEFR scales (e.g., admission to University, naturalization), it is particularly important to be sure that the scales actually mirror empirical learner language. Here, very little research has been conducted (cf. e.g. Alderson et al. 2006; Alderson 2007; Fulcher 2004; Hulstijn 2007; Hulstijn et al. 2010; Little 2007; Wisniewski 2013, 2014).

CEFR scale calibration is based on practitioners' beliefs about second language competence as expressed in ratings. However, it is not clear to what degrees ratings actually reflect scale contents (Arras 2010; Eckes 2008; Pollitt/Murray 1996; Vaughan 1991). No learner language

---

[1] http://www.coe.int/t/dg4/linguistic/cadre1_en.asp

analyses were carried through in the CEFR scaling process to support empirical validity. MERLIN aims at contributing to research regarding CEFR scale validity. Linguistic correlates to contents of central chapter 5 scales were operationalized and are searchable on the interface.

Natural Language Processing (NLP) research

The MERLIN corpus provides valuable data for the development and evaluation of natural language processing tools for learner language (Meurers 2012). The corpus and its meta-information on learners and ratings readily support research on automatic native language identification, enabling such research to go beyond the current English learner focus. In a similar vein, the corpus has already been used for research on automatic proficiency classification for German (Hancke 2013). The MERLIN corpus also provides richly annotated learner data for the development and adaptation of NLP tools and applications that assist language learners in improving their vocabulary usage, coherence, spelling and grammatical accuracy.

2   Methodology

2.1 Data collection

The MERLIN texts stem from the written production parts of CEFR-related, standardized high-quality tests from telc (Frankfurt/Main, Italian and German tests, www.telc.net) and ÚJOP (Prague, Czech tests, www.ujop.cuni.cz). These institutions are ALTE-audited (www.alte.org). The tasks were in use until 2013 and are now freely available on the platform. On this basis, a trilingual learner corpus was compiled that can be queried according to a variety of pre-determined aspects that were annotated manually and automatically (Glaznieks et al. 2014).

2.2 Transcription

The hand-written original learner texts were transcribed in an xml-based editor (xml mind©) by testing institutions (telc and ÚJOP). The transcribers followed transcription guidelines (available on the interface) and the reliability of the transcripts was checked, initially for a sample of 5% of the texts per CEFR level. As many transcription errors were detected, in the end almost all texts had to undergo a revision stage.
The transcription guidelines included tags (inline annotation) for basic textual features such as unreadable or ambiguous stretches of language, foreign language words, emoticons, images, paragraphs, copied words from the rubrics, or greeting formulae. The anonymization (names, places) was part of the transcription process and was carried through based on the

guidelines. Transcripts served as a basis for annotations (see below). The transcription guidelines are available on the MERLIN interface *(<<documentation>>,* in German only).

2.3 Re-ratings

In the original tests the MERLIN texts were extracted from, test-takers received a score that was then weighted in different ways according to the importance attributed to writing in that particular test, leading to an overall pass or fail mark. For MERLIN, the procedure was necessarily different: the aim was to have a direct relation of texts to CEFR Chapter 5 scales of communicative language competence. Therefore, all texts were re-rated independently by professional raters.

The reliability of the re-ratings was examined with the help of Classical Test Theory and a Multi-Facet Rasch analysis. The latter is a probabilistic statistical procedure often used in language testing which allows for a correction of rating tendencies (e.g., leniency/harshness) and makes it possible to arrive at a fair average rating for each text. The intra-rater and inter-rater reliability was generally very high in MERLIN, with some exceptions for Italian. Therefore, the whole re-rating process was repeated for Italian resulting in a satisfying rating quality. The details can be found in the Technical Report (see *<<documentation>>* section).

In MERLIN, a holistic scale (based on the CEFR scale for general linguistic range) was used together with an analytical rating grid (rating criteria: orthography | grammatical accuracy | vocabulary range | vocabulary control | coherence & cohesion | sociolinguistic appropriateness) ranging from A1 to C2. Both instruments can be downloaded from the MERLIN *<<documentation>>* section. The fair average is calculated based on the holistic scale. If you compile your own corpus based on CEFR levels, these are also based on the fair average ratings. On the interface, you can access a rating profile with the original ratings for these rating criteria, as well.

Please note that many test-takers took a test which then turned out to be either too difficult or not much of a challenge to them. Therefore, on the MERLIN platform a distinction is made between the *CEFR level of a test* which need not be identical to the *CEFR level(s) of the rating(s)* (see also table 9 below. Both can be searched for separately (e.g., you can search for learners of German who took a B1 test but received only an A2 rating).

2.4 Manual annotations

Annotation is one of the core aspects of the MERLIN project. MERLIN has two types of annotations: **'target hypotheses'** and annotations of **learner language features**. Where possible, **automatized procedures** (see 1.4) were used but most annotations were carried through **manually**.

The annotation was organized in 2 blocks:
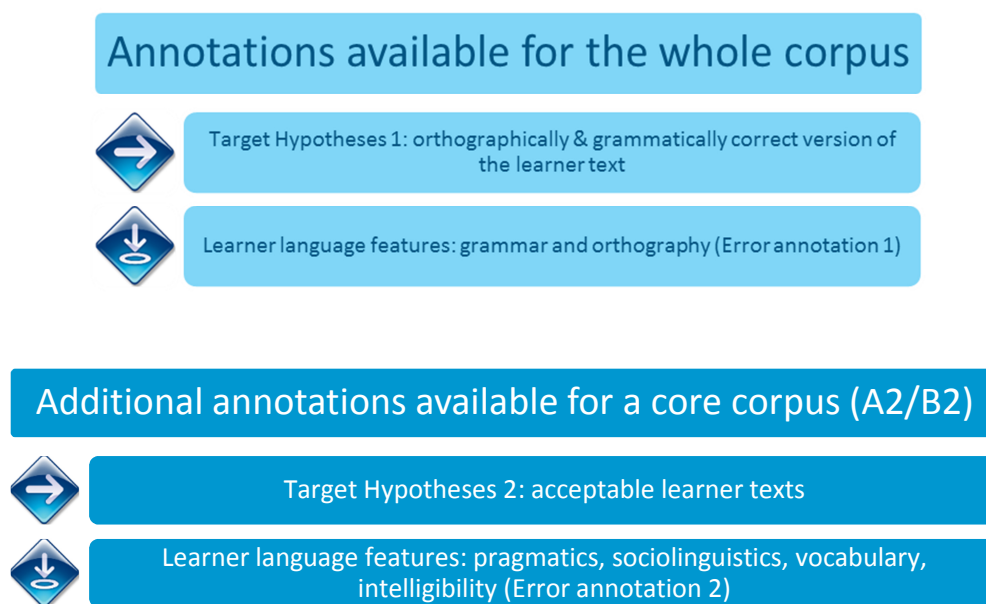
Graphic 1: MERLIN annotation architecture

2.4.1 Target Hypotheses (TH)

Research has shown that annotating learner language is a complex and partly speculative endeavor. Any mark-up of a learner language phenomenon requires a mental interpretation by the annotator. To guarantee transparency, coherence, and reliability of annotations, it is a good idea first explicitly write a 'target hypothesis' (TH), i.e. a corrected reconstruction of the learner text that a subsequent error annotation can build upon (Reznicek/Lüdeling et al. 2012). Also, target hypotheses are necessary for the successful implementation of many automated analyses (Díaz-Negrillo et al. 2010, Hirschmann et al. 2009). Thirdly, they can help future users of the MERLIN platform to understand annotations. MERLIN co-operates with the Falko project [2](Humboldt University, Berlin) which is one of the very few corpus initiatives that has a focus on target hypotheses and provides free access to the data. There are two types of TH1 (TH1 and TH2) in MERLIN which will be briefly explained in the following paragraphs.

Target hypothesis 1

Target hypotheses for orthographic and grammatical errors (TH1) were written for the complete MERLIN corpus. In TH1 writing, the annotator is asked to change the learner text as little as possible in order to create a grammatically and orthographically correct version of the original learner text ('minimal' TH). In this table, you find an example:

---

[2] https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko.

Learner text	Ich	habe	seit	5	Jahren	in	meinen	Heimatland	Deutsch	gelernt	(…)
TH1	Ich	habe	seit	5	Jahren	in	meinem	Heimatland	Deutsch	gelernt	(…)
TH1Diff							CHA				
Gloss TH1	*I*	*have*	*snce*	*5*	*years*	*in*	*my*	*home country*	*German*	*studied*	*(…)*

Table 2: Target hypothesis 1 (TH1), example

The following example by the same learner shows that in TH1, errors from other linguistic areas were ignored. There are content and technical reasons for this.

Learner text	Ich	habe		Srilankische	Aufenthalts	und	Reise	Spass	.
TH1	Ich	habe	einen	sri-lankischen	Aufenthalts-	und	Reisespass		.
TH1Diff			INS	CHA	CHA		MERGE		
Gloss TH1	*I*	*have*	*a*	*Sri Lankan*	*residence*	*and*	*travel fun*		*.*

Table 3: Phenomena not captured on TH1, example

While the orthographical (capitalization error, word boundary error, missing hyphen) and grammatical (missing article) errors are corrected in the TH1, the lexically erroneous form *Reisespass (instead of "Reisepass") was not substituted by another lexeme.

Target hypothesis 2

On a second level, extended target hypotheses (TH2) refer to aspects of sociolinguistic, lexical, and pragmatic deviations from what would normally be expected from a native speaker. TH2 thus aim at creating an *acceptable* version of the original learner text. For TH2, contextual aspects are taken into consideration. TH2 is an extension of TH1.

TH2 involve more subjectivity and difficulties in creating reliable decisions than TH1. This is a further reason to separate the two layers from each other. To illustrate the difference between TH1 and TH2, the following table might be useful:

Learner text	Ich	habe		Srilankische	Aufenthalts	und	Reise	Spass	.
TH1	Ich	habe	einen	sri-lankischen	Aufenthalts-	und	Reisespass		.
TH1Diff			INS	CHA	CHA		MERGE		
TH2	Ich	habe	einen	Sri-lankischen	Aufenthalts-	und	Reisepass		.
TH2Diff			INS	CHA	CHA		MERGE/CHA		
Gloss TH2	*I*	*have*	*a*	*Sri Lankan*	*residence*	*and*	*travel passport*		

Table 4: Target hypothesis 2 (TH2), example

Here, it becomes obvious that for the same sentence cited also for TH1, on the level of TH2 an additional lexical annotation is required ("Reisepass" [passport] instead of *Reisespass [travelling fun])

2.4.2 Annotation of learner language features

An important guiding principle in MERLIN is the view of learner language as a system in its own that cannot be satisfactorily described with deficit-oriented error tags alone. It is important to stress that MERLIN annotation do include many error tags, but that particularly on EA2, there are annotation tags that record phenomena which are not errors, e.g. the realization of the speech act REQUEST or formulaic sequences.

The MERLIN annotation scheme thus represents a selection of meaningful, valid, and feasible features ('tags') that are manually annotated and that are supported by the MERLIN Computational Linguists team. It contains the following features:

G_ Grammar

G_Agr	agreement (subject and verb)
G_Art	article
G_Clit	ITA: clitic
G_Conj	conjunction
G_Inflect_inexist	inexistent inflection (nouns, adj, verb)
G_Morphol_wrong	wrong inflection (nouns, pronouns, adj)
G_Neg_negdoub	CZE: double negation
G_Neg_neggen	negation general
G_POS	part of speech error
G_Prep	preposition
G_Refl_pronrefl	reflexive pronoun
G_Refl_pronreflposs	CZE: possessive reflexive pronoun
G_Valency_complnumb	verb valency: number of obligatory arguments
G_Verb_asp	verb: aspect (CZE+ITA)
G_Verb_compl	verb formation (morphol.)
G_Verb_main	main verb
G_Verb_md	verb: mood
G_Verb_tns	verb: tense
G_Verb_vc	verb: voice
G_Wo_womaincl	word order in main clause
G_Wo_wosubcl	word order in subordinate clause

O_ Orthography

O_Abbrev	abbreviation
O_Apostr	GER+ITA: apostrophe
O_Capit	capitalization

O_Graph_act	CZE+ITA: diacritical marks
O_Graph_graphgen	general grapheme error
O_Graph_trans	grapheme transposition
O_Punct	punctuation
O_Wordbd	word boundary

G_ Intelligibility

G_Intelltxt	intelligibility of text
G_Intells	intelligibility of sentence

V_ Vocabulary

V_FS	formulaic sequence
V_Sequence V_lexgrammerr_incompr	incomprehensible sequence caused by accumulation of lexical/grammatical error(s)
V_form_nonexist	non-existing form (word or formulaic sequence)
V_FS_form_incompr	formulaic sequence: limited intelligibility
V_semdenot	semantic error: denotation (word or formulaic sequence)
V_semcon_att	semantic error: connotation (attitude), (word or formulaic sequence)
V_Word_semimprec	semantic error: precision (word or formulaic sequence)
V_Wordform_deriv	word formation error: derivation
V_Wordform_comp	word formation error: composition
V_FS_form	formulaic sequence: form error

C_ Coherence/Cohesion

C_Con_accur	connector accuracy
C_Coh_jump	content jumps
C_Coh_ref	reference
C_Coh_txtstruct	metacommunicative device

S_ Sociolinguistic appropriateness

S_Txt_grfw	salutations/complimentary closes
S_Txt_opcl	opening/closing formulae
S_Form_gen	inappropriate style (formality)
S_Form_addr	inappropriate addressing (formality)
S_Var_clit	ITA: lexicalised clitics (verbi procomplementari)
S_Var_duppron	ITA: personal pronoun redundancy
S_Var_synstr	ITA: marked syntactic structures
S_Var_che	ITA: 'che polivalente'
S_Var_woweil	GER: main clause word order after 'weil'
S_Var_partik	GER: modal particles

P_ Pragmatics

P_Pol_dir	politeness - overly direct language form
P_Request_direct	direct REQUEST
P_Request_indirect	indirect REQUES

Table 5: Annotated learner language features (tags & definitions)

These tags were chosen from a comprehensive list of features and indicators that were gathered in the project. Possible annotations were **collected from**…:

1) the platform users' perspective (based on a user study, on textbook and language test analyses)
2) the CEFR (by operationalizing elements of chapter 5 scales, CoE 2001)
3) Second Language Acuiqistion research (based on an extensive review of research literature)
4) learner texts (based on an inductive analysis of 10% of all learner texts)

There are many sources of information users of MERLIN can access with regard to the annotations:

- the **bibliography** that was used for choosing meaningful annotation tags in MERLIN, in the reference section of this manual
- the **annotation scheme** with all annotation tags that were implemented in MERLIN after a practicality check of a maximum list of annotations (available for **download**)
- documents like the **MERLIN annotators' manual** (EN) and a **documentation of additional annotation issues** with fine-grained solutions to single annotation phenomena in the three project languages and other materials you can find on the interface (available on request).

- help functions on the interface, like a **list of annotation tags with examples**, **screenshots** explaining the annotation **tiers**, a **glossary**, and much more
- a **black book** that sums up experiences with annotations that future project in this field might want to work their way around (available for **download**)

Annotation tags suggested by practitioners

MERLIN is made for practitioners who work with the CEFR. Therefore, it is important to make the annotations as helpful for them as possible. To that end, a **user study** was carried through details of which are described in two reports available on the interface *(<<documentation>>)*. The first part focused on the usefulness of content aspects, while the second part took into consideration the technical side of the MERLIN platform usability. The user study delivered important information on the annotation tags to be included. The annotation scheme contains information regarding the tags that stem from the user study.

A further possibility to understand users' needs is the integration of aspects of L2 acquisition that are commonly treated in **textbooks and in language tests** into the MERLIN annotation scheme. In MERLIN, for example, "Tangram" for German (Dallapiazza 1998), "Rete!" for Italian (Mezzadri 2000) and "Brána jazyka českého otevřená" for Czech (Hasil 2007) were among the analyzed books. Also, the analysis of UJOP and telc language tests revealed certain notorious topics that delivered information for the annotation scheme.

Example annotations derived from these analyses include orthographical errors like erroneous capitalization, the incorrect use of the apostrophe in German and Italian, or grammatical errors such as the verbal aspect in Italian. In the lexical area, false friends or the use of idioms are recurrent topics that are mirrored in the MERLIN annotation scheme.

Learner texts as source for deriving annotation tags

10% of all learner texts in the corpus were analyzed by hand. This qualitative and inductive approach revealed a number of phenomena that were considered worthy of integration into the annotation scheme. Examples are word formation errors, many different types of problems related to the use of formulaic sequences, problems regarding the choice of the appropriate register in terms of formality or politeness. Interestingly, the majority of phenomena from this category belong to categories other than grammar and orthography.

Research-based annotation

A major effort went into an extensive research literature review for the different areas of language involved in MERLIN annotations. A full discussion of the single tags is not possible here, so we will try to give you a short overview. The references cited (and many more) can be found in the bibliography.

Annotations of **grammatical** phenomena include agreement, word order, negation, part-of-speech errors and much more. Thus, many accuracy and complexity measures can be calculated (see Lu 2010, 2011; Wolfe-Quintero et al. 1998, Ortega 2003,2012, Housen/Kuiken 2009). **Orthography** is an area that is not very extensively worked on in research. MERLIN annotation allows to access the relative/absolute frequency of different types of orthographic errors regarding, for example, punctuation, capitalization, or diacritics (Al-Jarf 2009, Bredel 2010, Cook 2005, Granger/Bestgen 2011, Perfetti/Rieben/Fayol 1997, Rimrott/Heift 2008, Sassoon 1995). These annotations are available for the whole corpus.

The MERLIN pilot core corpus annotations (EA2)

For a **small subcorpus**, aspects from other linguistic areas were annotated, too (EA2, error annotation 2). These subcorpus texts have obtained TH1 & EA1 and, in addition, TH2 and EA2. In the future, it would be desirable to have these **explorative pilot annotations** re-checked to then be able to cover the whole MERLIN database with TH2 & EA2 annotations. For now, we ask users to handle the core corpus annotations with caution.

Annotation tags for **vocabulary** take into consideration the manifold dimensions of lexical knowledge such as its accuracy, its depth, breadth, and sophistication (Nation 2001, 2007, Read 2000). Here, MERLIN has a strong focus on formulaic sequences (Wray 2002) which play a particular role in the acquisition of foreign languages (e.g, Pawley/Syder 1987, Schmitt et al. 2004). Many lexical tags regarding formulaic sequences are not error-related, but aim at capturing structures of particular interest in the acquisition process.

Sociolinguistic competence is defined as 'the capacity to recognize and produce socially appropriate speech in context' (Lyster 1994: 263). As for EA2 annotations in general, the annotation of phenomena pertaining to this competence (subjectivity, reliability) is methodologically challenging. The tags applied are in line with the project design and do not always comply with what is usually analyzed in the field (e.g., proficiency and amount of language contact, or diasystematic variation in learner texts as compared to L1 variation, Baker 2010, Baylea 2007, Bayley/Regan 2004, Biber/Finegan 1994, Hudson et al. 1005, Hymes 1974, Mougeaon/Dewaele 2004, Regan et al. 2009, Van Compernrolle/Williams 2012, Yu 2012, Zuskin 1992). In MERLIN, the appropriateness of language forms with regard to 'formality' is annotated (e.g., substandard forms, use of forms that pertain to oral communication, overly formal language, see Koch/Oesterreicher 2011), and tags that are text-type tailored such as greetings or opening and closing formulae are integrated. Single language-specific variational aspects chosen reflect choices of structures that either do not pertain to the written language and/or seem to be on their way to be accepted as standard variants, but would not normally be accepted in the task types they can be found in in the MERLIN texts.

In the area of **pragmatics**, **coherence/cohesion** is annotated mainly by a mark-up of the use of metacommunicative devices, also independently of correctness, connectors, and the annotation of reference problems (Bachmann 2002, Halliday/Hasan 1976, 1989, Castro 2004, Carlsen 2010, Chiang 2003, Cornish 2009, Fabricius-Hansen 2005, Louwerse/Graesser

2004, McNamara et al. 1996, McNamara/Kintsch 1996, Spooren/Sanders 2008). Also, the speech act of REQUESTING is annotated (Al-Gahtani/Roever 2012, Bardovi-Harlig 2013, Barron 2003, Blum-Kulka 1987, 1991, Blum-Kulka/Olshtain 1984, Cho 2005, Held 1995, Nuzzo 2007, Trosborg 1995, Veddersen 2007).

CEFR-based annotation

To find out if the CEFR scales reflect learner language, it is important to operationalize their descriptors without making use of human ratings which have often turned out to not be based on rating instruments even when they are reliable (Eckes 2008, Wisniewski 2010). If scale descriptors are put into a measurable form like this, the relationship between selected CEFR scales and learner language becomes much clearer.

In this operationalization process, exceedingly vague, self-referential, or subjective terms in the level descriptions had to be excluded (e.g. "Can sustain relationships with native speakers without […] requiring them to behave differently than they would with a native speaker", sociolinguistic appropriateness, B2, CoE 2001: 122, but also aspects that were clearly related to spoken language only were ignored (e.g. "Can…keep up group discussions […]", sociolinguistic appropriateness scale, B2, CoE 2001: 122) (cf. Wisniewski 2013, 2014). If, however, a level description mentions "greetings", "content jumps", "intelligibility", "idiomatic expressions" or "phrases" as characteristics of specific CEFR levels, these features were checked for feasibility in the MERLIN annotation scheme, even if these so-called "scale variables" might not play a role in research or are often hard to clearly define. These annotations allow to check the empirical relevance of the CEFR scales involved. It would be a sign of empirical validity if the scale contents were sufficiently salient and reliably observable in learner performances. MERLIN cannot offer a complete validation of CEFR scales, but it focuses on a selection of meaningful aspects.

Accessing annotations on the interface

Annotations can be accessed **directly** via the search functions of the interface *(<<Advanced search>>, <<Define a subcorpus>>*), where every single occurrence is displayed in context.

Another possibility to access annotations is to use them for **statistical measures**. The simplest possibility is to count the total number of annotation tags occurring in a (sub)corpus. In many cases, though, it is more meaningful to use annotations for the calculation of normalized measures (i.e., in MERLIN, per sentence, T-unit, or token). To give you an example, this allows you to compare the average number of morphological errors per sentence in Czech B1 vs Czech B2 learner texts. Normalized measures of manually annotated phenomena are calculated on the basis of automatized segmentation procedures (see below). The frequency-based measures are available in the *<<statistics>>* section on the interface, along with a number of more complex measures of complexity.

2.5 Automatized annotations

The automatic annotation in the MERLIN corpus serves to support the manual annotation and to make accessible a wide range of linguistic features for the calculation of indicators and in direct corpus searches.

The automatic annotations for the MERLIN corpus can be divided into four categories:

1. Linguistic units needed for the manual annotation:
    a) tokens
    b) sentences
2. Linguistic units needed for the calculation of measures:
    a) t-units
    b) a range of clause types
3. Linguistic annotation using existing definitions and tools
    a) part-of-speech
    b) lemma
    c) constituency and dependency parses
4. Linguistic annotation with MERLIN-specific definitions and tools
    a) repetitions within texts
    b) citations of task material

We have applied existing automatic annotation tools developed for the target languages in order to expand the range of available linguistic annotation beyond what would have been possible with time-consuming and expensive manual annotation. However, it is important to keep in mind that automatic annotation is particularly challenging for learner language, since learner language often deviates considerably from the target language across all levels of linguistic analysis, from spelling to semantics.

The following tools were used for all three MERLIN languages:

Texts were tokenized using the tokenizer for Indo-European languages from LingPipe and the resulting tokenization was then corrected by hand. Sentences were annotated with the OpenNLP sentence segmenter. Repetitions were identified using the Saphre library on the basis of the automatic part-of-speech and lemma annotation.

Please refer to the *<<MERLIN for research>>* section to learn more about the language-specific tools used for automatic annotation.

2.6 Quality control

In order to organize and control annotation reliability, a number of measures were taken. All instruments (TH 1 & TH2 rules, annotation scheme for EA1 and EA2) were piloted and revised before their implementation. Piloting was organized in two steps. First, the

annotation procedure itself was tested to get a first idea of how to concretize and change the annotation guidelines and instruments. Then, in a piloting process, a restricted number of texts were annotated by all annotators of a language team in order to again check the practicality and the sufficiency of the annotation guidelines and in order to detect possible technical problems.

Secondly, all annotations are based on guidelines (annotator manual, see interface). The guidelines are enriched by fine-grained decisions on single aspects of annotation (document on additional annotation issues, see interface). Thirdly, the reliability of the annotations is controlled. Reliability of annotations was controlled for 5% of the texts on each test level for target hypotheses and error annotation. Different methods were applied:

In a qualitative approach, half of the files are annotated independently by the coders to then be commonly discussed with the aim to arrive at a consensus. These texts served as a reference throughout the annotation process. The qualitative approach turned out to be extremely important for a common understanding of the annotation scheme. In a double-blind procedure, the second half of the files checked for reliability was annotated by all coders without their knowledge. The annotations in these files were checked for coder reliability qualitatively and quantitatively.

PART II: User guide

3. The MERLIN documentation section

The macrostructure of the freely accessible MERLIN platform is organized in a **documentation area** (vertically placed on the left of the interface, see graphic 2) and a **search area** (horizontal search bars, see graphic 23). Chapter 3 explains the most important contents and functions of the documentation section.
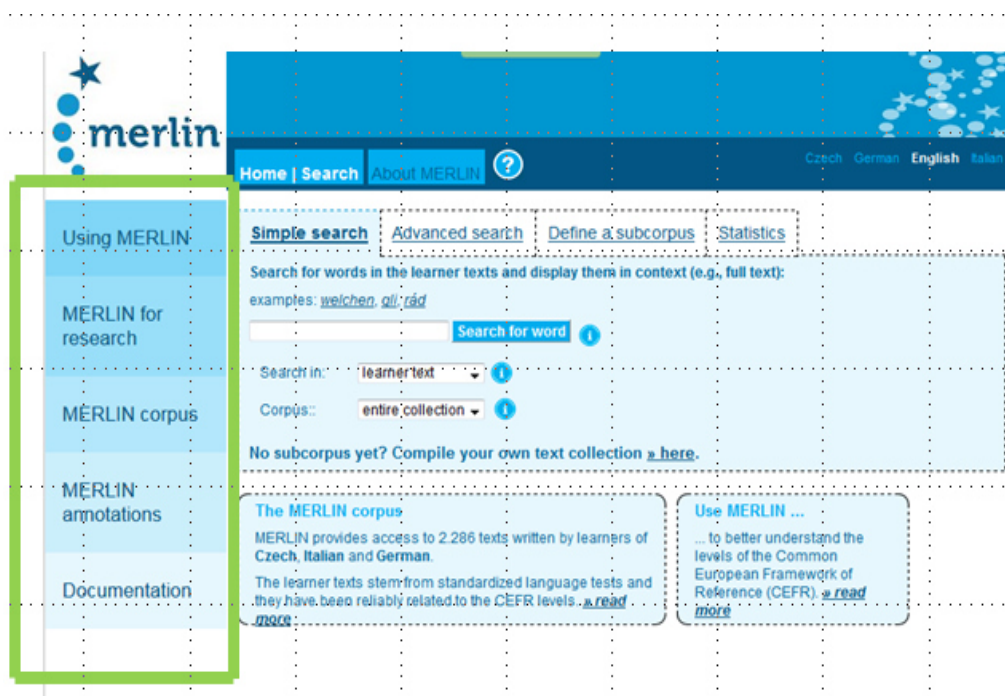
Graphic 2: <<documentation>>

3.1 <<Using MERLIN>>
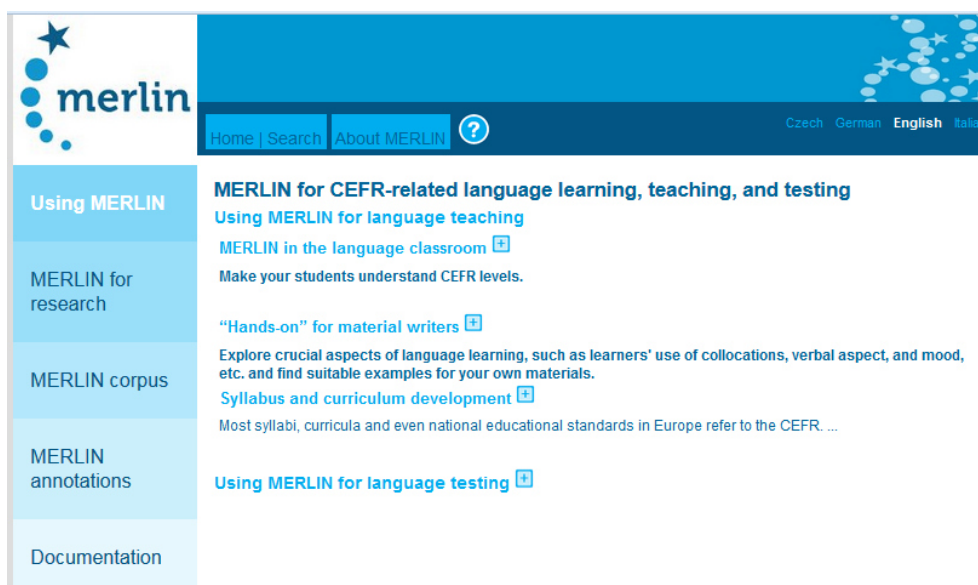
Graphic 3: <<using MERLIN>>

In the *<<using MERLIN>>* section, you get general advice on possibilities for applying MERLIN in different professional settings. This section is useful to give you a first idea of what the project is all about.

You can unfold the chapters (click ⊞) to learn more about, for example, how you can use MERLIN for language teaching or for developing teaching materials.
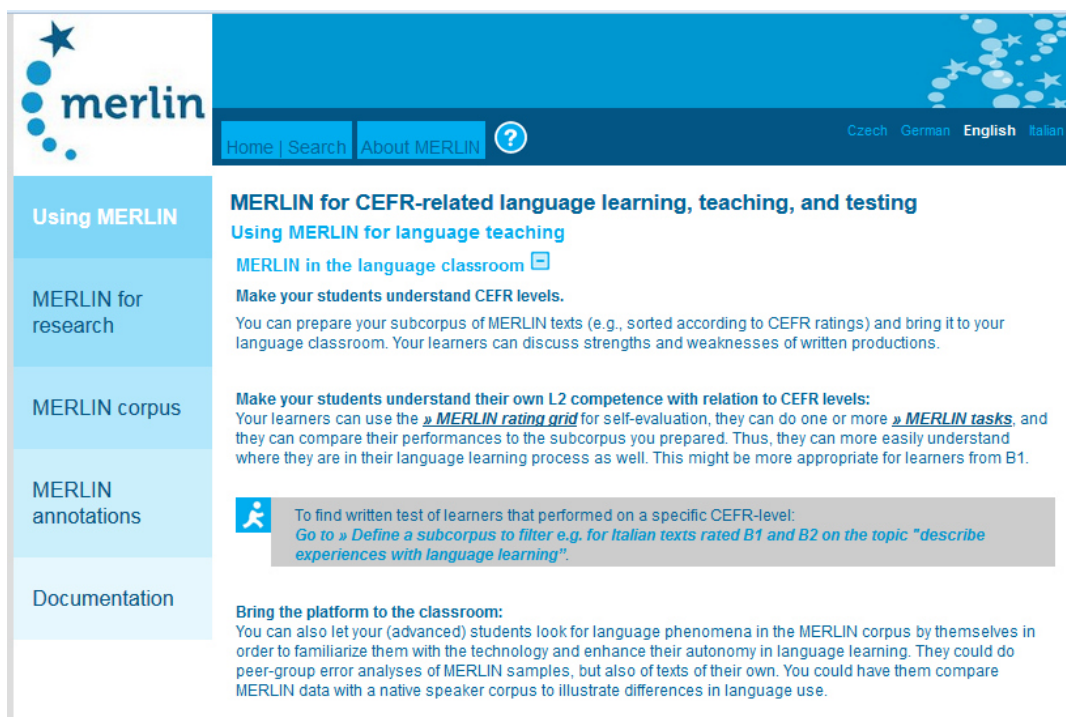
Graphic 4: <<using MERLIN>>, <<Using MERLIN for language teaching>>

3.2 <<MERLIN for research>>

The section *<<MERLIN for research>>* is meant for anyone interested in background information regarding different aspects of the project. Whenever you browse through the interface and would like to learn more about the project rationale, it is advisable to consult the *<<MERLIN for research>>* section.

Graphic 5: <<MERLIN for research>>

As graphic 5 shows, there is a wealth of information regarding the **linking of MERLIN texts to the CEFR.** All relevant documents like the rating grids, the tasks, the technical report regarding the quality of the ratings are accessible from here.

Also, the **workflow** that the MERLIN **data** underwent is outlined in this section: you can find out how the transcription and the annotation were carried through, with the help of which tools and schemes. You get information on quality control aspects of the manual and automatic annotations as well.

Thirdly, there is information regarding the possibilities to **use MERLIN in researching** the validity of the CEFR scales, second language acquisition, and Natural Language Processing of learner language. You also find a **list of relevant references**.

3.3 <<MERLIN corpus>>

The MERLIN corpus

The MERLIN corpus contains 2,286 texts for learners of Italian, German and Czech that were taken from written examinations of acknowledged test institutions. The exams aim to test knowledge across the levels A1-C1 of the Common European Framework of Reference (CEFR).

Texts and test institutions ⊞

Standardised texts used in written exams within the Common European Framework of Reference for Languages (CEFR) were extracted for the learner corpus to create texts for written assessments.

The relation to the Framework of Reference - the MERLIN rating grid ⊞

To ensure an immediate relation to the CEFR, specially trained testers re-rated all exam texts using the MERLIN rating grid that was developed within the project. ...

Test tasks ⊞

We provide a comprehensive overview of the test tasks by target language and CEFR level tested. ...

Available metadata ⊞

The MERLIN corpus in figures ⊞

Graphic 6: << MERLIN corpus>>

The section *<<MERLIN corpus>>* yields much information on the MERLIN data. You can learn more about the texts and the testing institutions, you can download the rating grids that were used, and you can see a list of the test tasks:

Test tasks ⊟

We provide a comprehensive overview of the test tasks by target language and CEFR level tested. ...

The level of the test may differ from the level that the learner text received in the re-ratings.

The tasks are represented using a *» grid* that was developed for these purposes by ALTE (Association of Language Testers in Europe, *» www.alte.org*). The grid contains detailed information about the tasks and the specific characteristics of the intended text, e.g. regarding topic, register, domain (author: Olaf Bärenfänger).

» General notes on task descriptions 📕

German

A1
Informal e-mail: ask a friend for help with finding an apartment 📕
Informal e-mail: arrange an appointment with a friend to go swimming together 📕
Informal letter: congratulate to birth of a child 📕

A2
Formal letter to housing office 📕
Informal letter: ask friend to take care of pet 📕
Informal letter: offer a ticket not used to a friend 📕

B1
Informal letter for New Year to a friend 📕
Informal letter to a friend announcing a visit 📕
Informal letter: birthday congratulations 📕

B2
Formal letter: ask for information at Au pair Agency 📕
Formal letter: Au pair writes letter of complaint to Agency 📕
Formal letter: apply for internship in sales department 📕

C1
Essay: why it's of value to learn German 📕
Online article: about sticking to one's traditions and "assimilation" in a new environment 📕
Report: about the housing situation 📕

Graphic 7: MERLIN test tasks for German (<<MERLIN corpus>>)

If you click on a task a pdf will open in an extra window that contains the task itself along with a detailed task description that is based on a Grid developed by the Association of Language Testing in Europe (ALTE, www.alte.org). The task description tells you more about the length of the task, the type of language in the expected response, its difficulty and much more.

i) Task input/prompt		
17	Language of input/prompt	German
18	CEFR level of input/prompt	A1
19	Time permitted or suggested for this task	n.a. minutes
20	Control/guidance	Semi-controlled
21	Content	Specified
22	Genre	Letter
23	Rhetorical function(s) of input	Describing, instructing
24	Imagined audience	Friend
25	Mode of input/prompt	Written
26	Topic or theme of input	Relations with other people
27	Integration of skills for input	Reading

Graphic 8: Detail of a task description (<<MERLIN corpus>>)

The *<< MERLIN corpus>>* section also provides information on available metadata (like age, gender, or mother tongue), and there is a table with the total numbers of texts available per test level and per rated overall CEFR level:

The MERLIN corpus in figures ▱

Number of texts per CEFR level of the test (test level) compared to the number of texts per CEFR level assigned in the re-rating (fair average)

	Test Level		Fair Average	
			A1	1
	A2	111	A2	189
Czech	B1	143	B1	165
	B2	188	B2	81
			C1	2
	A1	207	A1	29
	A2	202	A2	378
Italian	B1	201	B1	394
	B2	201	B2	2
	A1	206	A1	57
	A2	209	A2	297
German	B1	210	B1	331
	B2	204	B2	293
	C1	204	C1	42
			C2	4
Total	2286		2265	

Graphic 9: Number of texts taken on the different CEFR levels & number of ratings on each CEFR level (<<MERLIN corpus>>)

And a table with information regarding the depth of annotation of the MERLIN texts:

	Czech	German	Italian
Texts	442	1033	813
TH1	440	1033	813
EA1	361	752	754
TH2	231	275	154
EA2	198	258	85

Graphic 10: The MERLIN corpus in figures: Number of texts with target hypotheses 1 &2 and error annotation 1 & 2 (<<MERLIN corpus>>)

3.4 <<MERLIN annotations>>
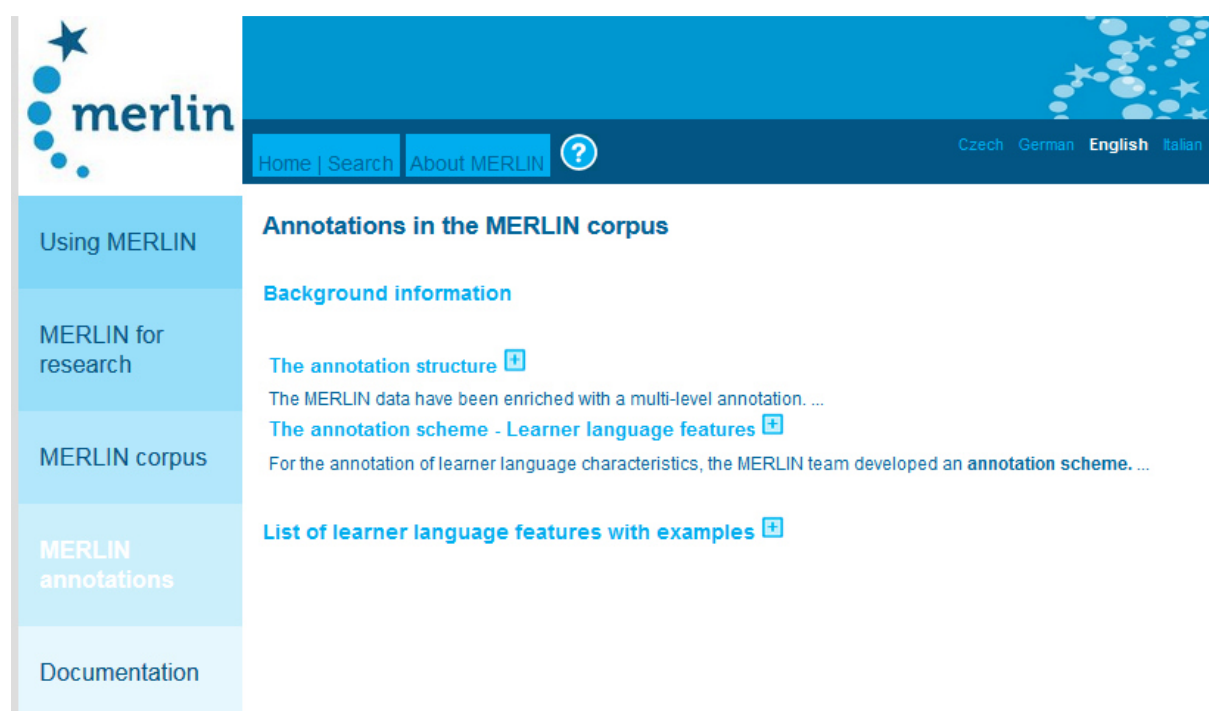
Graphic 11:   << MERLIN annotations>>

This section contains information regarding MERLIN annotations, with a focus on the manual annotations.

You can get an overview of the annotation architecture:

Graphic 12: MERLIN annotation structure, full corpus (<<MERLIN annotations>>)

Also, you get access to the **annotation scheme** all manual MERLIN annotations are based upon, with examples for all three languages, a glossary for less common terminology, and clear definitions of each tag. Furthermore, you can download a comprehensive **progress documentation** of difficult questions that arose regarding single annotation aspects. If you come across an annotation that seems questionable to you, this document might be useful (it is also accessible from *<<documentation>>*).

If you do not want to read through the full tag definitions, instead of downloading the complete annotation scheme, you can also consult a list with all annotation tags and examples:

List of learner language features with examples ⊟

Grammar | Orthography | Intelligibility | Vocabulary | Cohesion/Coherence | Sociolinguistic appropriateness | Pragmatics

GRAMMAR TAGS	Example*
word order in main clause	*[Vielleicht du könntest mir bei meine Wohnungssuche helfen.] *[Sollst du Wasser und Bikini mitbringen.]
word order in subordinate clause	*[wenn haben Sie Zeit,] dann bitte sagen Sie mir.
negation general	*Ich habe [nicht] Zeit.; *Er wird dort arbeiten [nein].
CZE: double negation	*[mám] žádný čas {nemám žádný čas}; *nikdo [volal] {nikdo nevolal}
verb valency: number of obligatory arguments	CZE: *Petr vstává v 6 hodin. On nesnídá, protože [on] nemá hlad. GER: *Er hat uns nicht gesagt, ob {er} kommen will.
agreement (subject and verb)	*Jana [hast] gelesen, *Jana [sind] müde
reflexive pronoun	CZE: *smála [si]; GER: *er [entschuldigt], *Laura und Ferdinand reden [sich] ITA: *[se] {si} lava ogni mattina
CZE: possessive reflexive pronoun	*potřebuju [moji] knihu, * vidím [mého] otce
inexistent inflection (nouns, adj, verb)	adjective: *ein [blaus] Himmel {blauer}; [teuerer] {teurer}; [größen] {großen / größeren} noun: *das schöne [Hause], *[euche] [Fahrrade]

Graphic 13: List of annotation tags with examples (<<MERLIN annotations>>)

3.5 <<Documentation>>

Graphic 14: <<Documentation>>

Here, all freely available MERLIN-related documents can be found, i.e.:

- reports produced in the project, e.g. on the quality of the ratings and the usability of the platform (user studies)
- all test tasks used
- grids used for rating;
- guidelines and schemes used for transcription and annotation (e.g., annotation scheme)
- publications, presentations and the like by the MERLIN team
  …

3.6 <<Download whole corpus>>

Download MERLIN corpus

MERLIN offers free download of all learner texts included in the corpus.

(work in progress: to come soon)
Download all learner texts including **metadata** (information about the learner, the test, and ratings):

CZECH	zipped PDF files	zipped TXT files
GERMAN	zipped PDF files	zipped TXT files
ITALIAN	zipped PDF files	zipped TXT files

Download all learner texts including **metadata** AND **target hypotheses**:

CZECH	zipped PDF files	zipped TXT files
GERMAN	zipped PDF files	zipped TXT files
ITALIAN	zipped PDF files	zipped TXT files

Please note: If you want to define your own subset of learner texts (subcorpus), go to **define a subcorpus**. From there, you can download a sample or all texts from your subcorpus as a single multi-page file (TXT or PDF).

Graphic 15: <<Download corpus>>

This section gives you an immediate download option for the whole MERLIN corpus. You can download zipped files for Czech, Italian, and German in .txt or .pdf. You can decide whether you want to download the original learner texts with metadata (e.g., L1, age) or whether you also want to include target hypotheses.

If you are interested in a more specific collection of texts, go to *<<define a subcorpus>>* first. There, you can create your customized subcorpus, download it, search in it *(<< Simple/Advanced search>>)* or have statistical measures displayed *(<<statistics>>).*

4. The MERLIN search functionalities
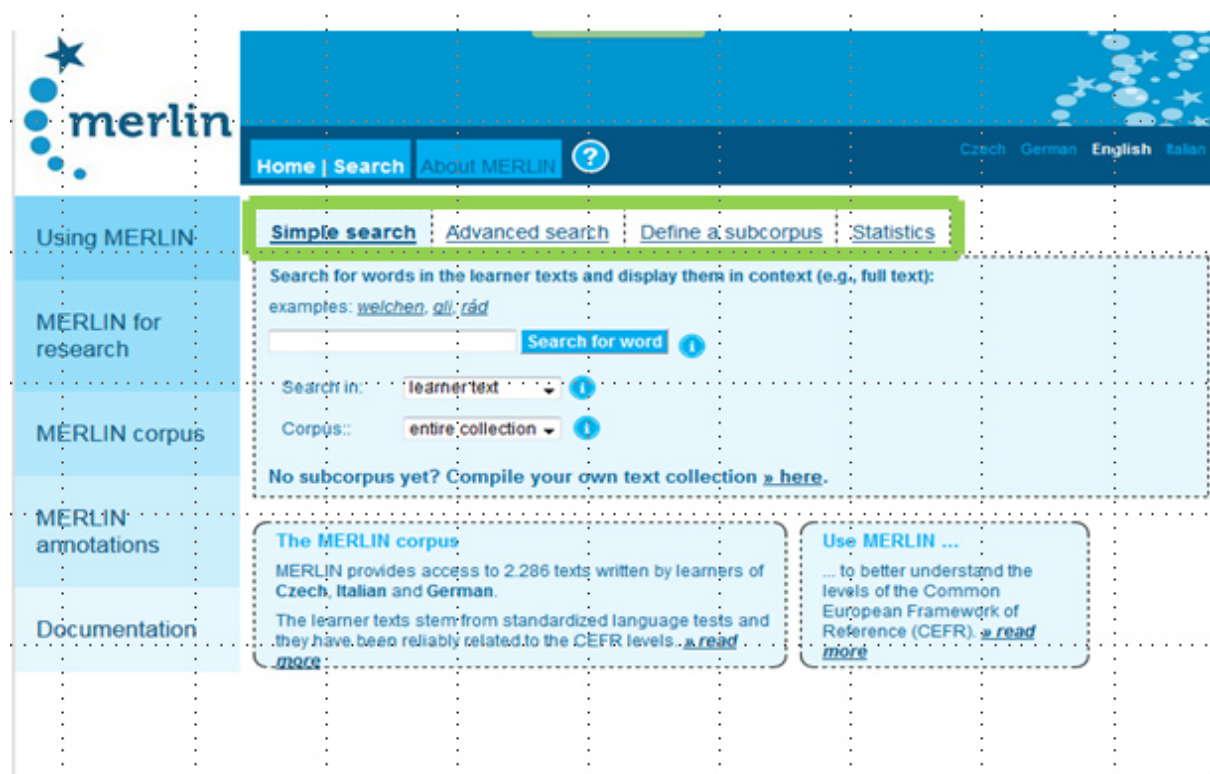
Graphic 16: search area

The search options include:
- a simple search
- an advanced search
- a "define a subcorpus" section
- a statistics section

4.1 <<Define a subcorpus>>

It is recommendable to start using the search option by defining a collection of texts one is interested in *(<<Define a subcorpus>>,* see graphic 17:

Graphic 17: Search interface <<Define a subcorpus>>

Here, it is possible to specify criteria according to which a collection of learner texts (a so-called subcorpus) can be tailored to users' needs. The criteria are:
- target language of the text
- original CEFR level of the test
- fair average CEFR rating of the texts (see section 2)
- one or more CEFR levels of single rating criteria (grammatical accuracy | vocabulary range | vocabulary control| sociolinguistic appropriateness | coherence & cohesion | orthography)
- test task

Furthermore, it is possible to sort text according to learner information:
- mother tongue (L1)
- age
- gender

Also, texts can be selected according to up to three learner language features and/or words:

Graphic 18: Specifying characteristics of a subcorpus, learner language features(<<Define a subcorpus>>)

The subcorpus needs to be given a name that can be chosen by you; by clicking on "define subcorpus and show texts", this subcorpus will be available for further searches for 24 hours. After that, you will have to redefine the subcorpus.

<<Define a subcorpus>> output

By clicking on "define subcorpus and show texts", a result similar to the one displayed in graphic 19 will be visible.

Graphic 19: Exemplary output of <<Define a subcorpus>>

In this output, you get a list of all the texts that match the criteria used for defining the subcorpus. In this case, the user gathered all texts that were written in Italian with regard to a specific MERLIN task (not in the screenshot).

The output page specifies the number of texts found ("total hits: 75"). From the ID in the first column, you can understand the target language ("Italian"), the overall rating (differs in the example), and the L1 (in the example: French and Polish).

You can **download** the documents (or only a selection of the subcorpus texts) with or without metadata and with or without target hypotheses in different formats by clicking "download documents".

Also, you can click on "**View learner text and TH**" to get the original text and the target hypothesis 1 or 1 and 2 (an example is shown in graphic 20).

Graphic 20: exemplary output of "view learner text and TH" in <<Define a subcorpus>>output

By clicking on "**View learner info and ratings**", metadata for a specific text is displayed (graphic 21 shows the metadata pertaining to the text in graphic 20 above):

Graphic 21: exemplary output of "view learner info and ratings" in <<Define a subcorpus>> output

The subcorpus can be further explored in the Simple or the Advanced search or in the statistics section.

33

4.2 <<Simple search>>

In the *<<Simple search>>,* it is possible to search for word forms:

Graphic 22: Search interface <<Simple search>>

The search can be run …
- in the learner text or in the target hypotheses (TH1 or TH2),
- in the entire MERLIN corpus or in a subcorpus that you specified *(<<Define a subcorpus>>)*

The simple search is not lemma-based (all word forms of a lexical entry):it only refers to the exact word form you enter. Thus, if you enter, e.g., "abholen" (like in the example below, graphic 23), you will not get results for "abholst" or "hole…ab".
You can use the *virtual keyboard*, if needed. Please also be aware of the fact that MERLIN is based on a limited number of tasks which elicit a constrained range of vocabulary when using the simple search.

<<Simple search>> output
The simple search output gives you the word you looked for in its immediate context (so-called keyword in context, or KWIC):

Graphic 23: Exemplary output of <<Simple search>>

By clicking on the key word in context, the full learner text will be displayed. By clicking on *<<view learner info and ratings>>*, the metadata of the text will be shown. You can also directly download a .pdf file of the task from there and copy the author's ID if you want to run more specific searches:

Graphic 24: Exemplary output of <<view learner info and ratings>> in <<Simple search>> output

4.3 <<Advanced search>>

This section allows you a more sophisticated access to the MERLIN corpus. It is possible to combine the search for lemmas/words with the search for a variety of annotations that are available in MERLIN.

You can search…
- in the learner text or in the target hypotheses (TH1 or TH2),
- in the entire MERLIN corpus or in a subcorpus that you specified (*<<Define a subcorpus>>)*

Graphic 25: Search interface <<Advanced search>>

Furthermore, you can decide whether you want to search for 1 or 2 words or lemmas that …
- are directly **adjacent** or appear with a specified number of words in between
- belong to specific **word class** you are interested in (based on automatic part-of-speech (POS) annotation). If you click on the dropdown menu after having selected the target language, you will get a list with POS abbreviations and short explanations (see graphic 26)
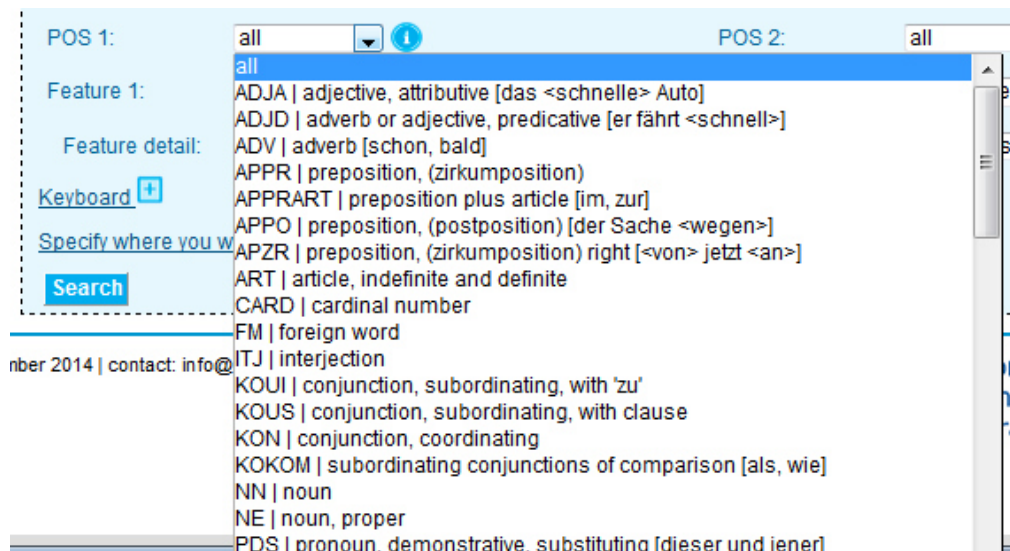
Graphic 26: Defining POS tags, <<Advanced search>> (detail, German)

Also, you get access to the **manual annotations** available in MERLIN. To that purpose, choose
- *<<feature 1>>* first to specify what category of annotation you are interested in, e.g., grammar, vocabulary, orthography
- As an option, you can further narrow down the search by defining *<<feature detail>>*: Here, you get a list with all **annotation tags** that have been used in a certain annotation category (in the graphic below, you find an example for grammar). If you do not choose a specific tag here, all grammar tags will be displayed if you chose "grammar" in the *<<feature 1>>.*
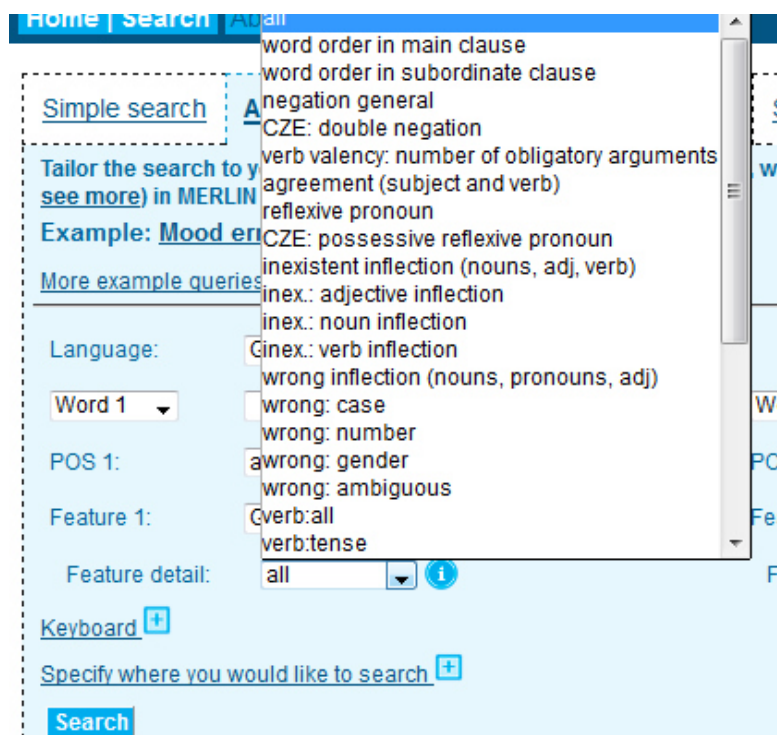
Graphic 27: Defining <<feature details>> in an <<Advanced search>> (detail, grammar)

You can execute this procedure for one word/lemma or for a combination of two words/lemmas. Please note that it is not possible to search for learner language features without specifying a word or lemma first.

If you need more information regarding the annotated features, in the *<<MERLIN annotations>>* section you can consult a list with all tags and examples, learn more about the MERLIN annotation architecture, and download the MERLIN annotation scheme.

<<Advanced search>> output

The *<<Advanced search>>* section uses the open source search and visualization architecture ANNIS (www.annis-tools.org) which is why its output looks different from what you get in the other MERLIN interface output sections. The following screenshots guide you through the output.

In graphic 28, the lemma "gebären" ("to give birth to someone") is entered in the *<<Advanced search>>.*

Graphic 28: Exemplary lemma search, <<Advanced search>>

In the output, you find
  (1)   the number of hits and the number of documents with hits (left hand side)
  (2)   the corpus you searched in (bottom left)
  (3)   a button that removes the left hand side of the output which is important to get to a full view of the learner texts

(4)   the possibility to get back to the *<<Advanced search>>* - your search will not have
be saved there

(5)   a link to more information *(<<About this search output>>);* a new window with
information regarding the single tiers of the annotation will open (see chapter 5
of this document)

(6)   a link to a list with short explanations of all abbreviations, mainly annotation tags,
that are used in the output will open in an extra window *(<<abbreviations>>)*

(7)   the MERLIN help function
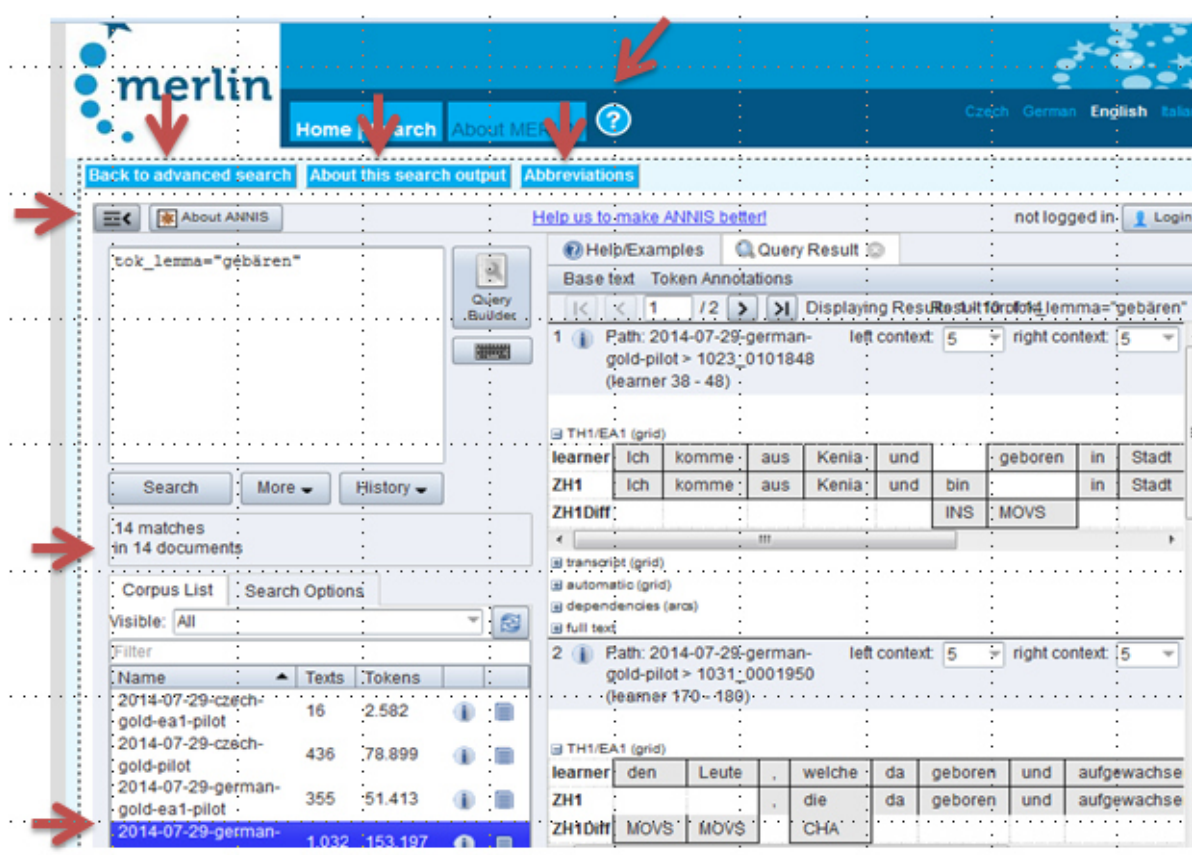
Graphic 29: Exemplary lemma search output, <<Advanced search>>

If you hide the left vertical output area which gives you meta information by clicking on

, it is easier to look at the learner texts directly:

Graphic 30: Exemplary lemma search output, hidden meta information on left side, <<Advanced search>>

You see the lemma searched for in its immediate context which you can enlarge to up to 25 tokens on either side:

Graphic 31: Modify amount of context of a lemma in <<Advanced search>> output

Please note that you can access automatic annotations, a full view, the transcript, and dependency arcs of the search result. In graphic 32, there is an example in which the automatic annotation is shown in a grid:

merlin

Home | Search   About MERLIN

Czech German **English** Italian

Back to advanced search | About this search output | Abbreviations

About ANNIS   Help us to make ANNIS better!   not logged in   Login

Help/Examples   Query Result

Base text   Token Annotations

|< < 1 /2 > >| Displaying Results 1 - 10 of 14   Result for: tok_lemma="gebären"

TH/EA (grid)

learner	den	Leute	,	welche	da	geboren	und	aufgewachsen	sind	.	Das
TH1			,	die	da	geboren	und	aufgewachsen	sind	.	Das
TH1Diff	MOVS	MOVS		CHA							

transcript (grid)

automatic (grid)

learner	den	Leute	,	welche	da	geboren	und	aufgewachsen	sind	.	Das
tok_lemma	d	Leute	,	welch	da	gebären	und	aufwachsen	sein	.	d
tok_lemma_bohnet	der	Leute	--	welcher	da	gebären	und	aufwachsen	sein	--	der
tok_pos	ART	NN	$,	PRELS	ADV	VVPP	KON	VVPP	VAFIN	$.	PDS
tok_pos_bohnet	ART	NN	$,	PRELS	ADV	VVPP	KON	VVPP	VAFIN	$.	PDS
tok_pos_stanford	ART	NN	$,	PRELS-SB	ADV	VVPP	KON	VVPP	VAFIN	$.	PDS-SB
sentence	Ich denke, wir müssen respektieren die Traditione den Leute, welche da geboren und aufgewachsen sind.										Das macht edes Land besonderes atraktiv.
repetition		id:4 count:2			id:10 count:2						
tunit				welche da geboren und aufgewachsen sind							Das macht edes Land besonderes atraktiv.

dependencies (arcs)

full text

Graphic 32: View automatic annotations in <<Advanced search>> output

The lines of the output table view are explained in *<<about this search output>>*.

4.4 <<Statistics>>

In this section, some fundamental statistical information is available. You can search in all MERLIN texts or in your choice of texts ("subcorpus", see section *<<Define a subcorpus>>.*)
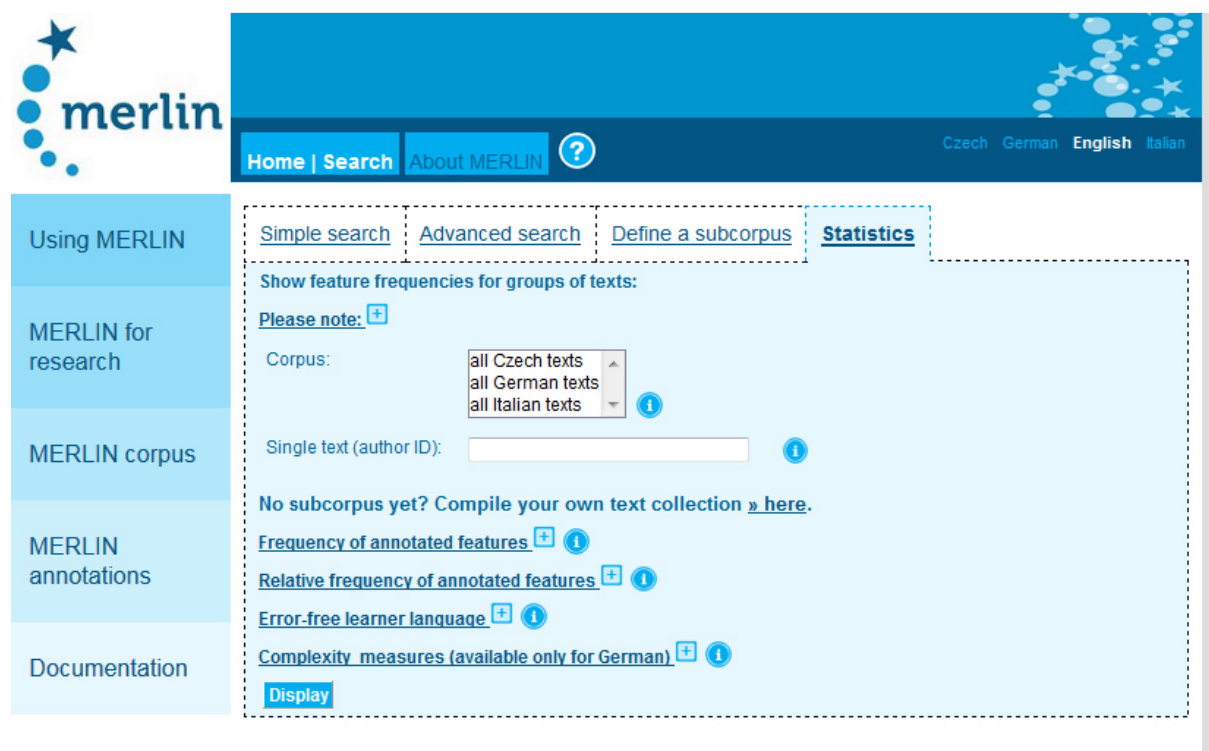
Graphic 33: Search interface <<Statistics>>

Notes of caution for interpretation of statistical measures
The MERLIN statistical measures are to be interpreted with extreme caution. They can deliver indications and tendencies, but they must not be over-interpreted as evidence for language learning routes. There are different reasons for this, some of which are listed here:

1) MERLIN is a small corpus – simple generalizations are not possible
2) Database for EA1 is different from database for EA2; the latter is extremely small
3) Contrastive analyses should be handled with care. It is not straightforward to compare aggregated measures of different languages because there are some language-dependent tags such as, for example, errors regarding aspect for Czech and Italian, but nor for German.
4) The measures strongly depend on the range of tasks used in MERLIN; this is true especially for the vocabulary and the other EA2 fields

In the *<<statistics>>* section, there are **four types of measures** available. For any choice you make, you can select multiple annotated features by holding down STRG:

(1)   *frequency of annotated features*

Absolute number of annotated features in your database. You can choose one or more feature categories (e.g., grammar, orthography). Then, you can either look for a count of all annotations regarding that category/those categories (e.g., all grammar annotations in German texts) or specify which annotations you are interested in on a more specific level (e.g., all morphological errors in German texts):

Graphic 34: Exemplary calculation of absolute frequency (all grammatical errors, German), <<Statistics>>

The **output** gives you an overview of the total number of the feature(s) searched for:
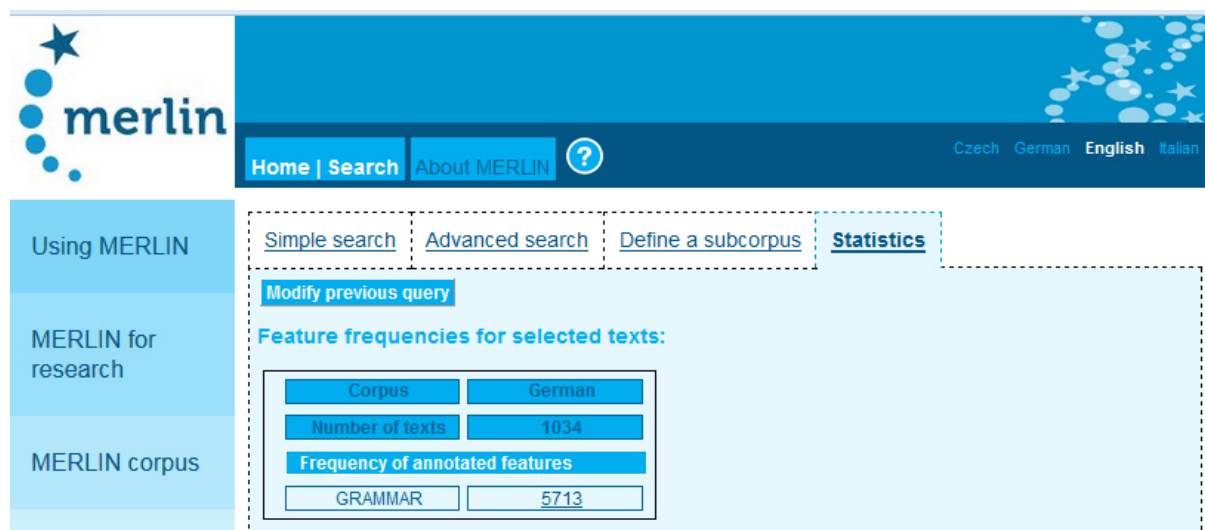
Graphic 35: Output of exemplary calculation of absolute frequency (all grammatical errors, German), <<Statistics>>

If you click on the number of hits (in our example, 5713), you will be directed to the <<*Advanced search*>> section with direct access to all annotations relevant to your search:
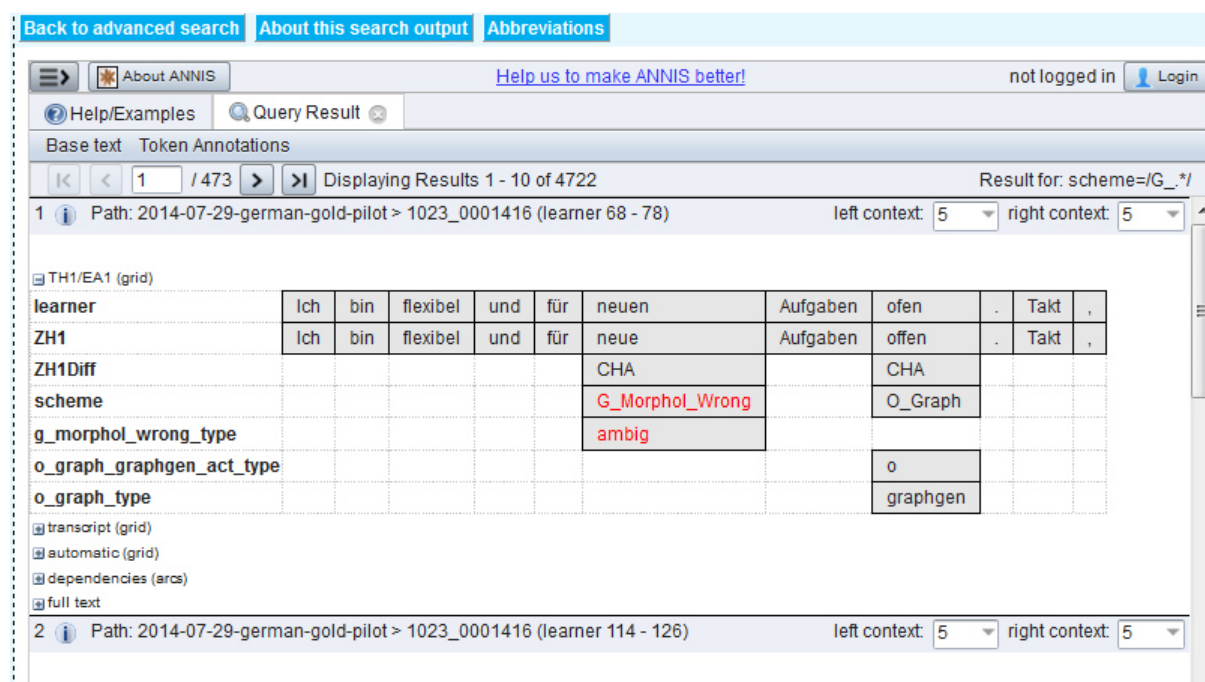
Graphic 36: All instances of exemplary frequency calculation (graphics 34-35) in <<Advanced search>> output

(2)   relative frequency of annotated features

Relative number of annotated features in your database per sentence or per token. As text length varies considerably throughout the corpus, it is useful to use normalized frequencies, i.e. numbers of occurrences of one or more phenomena with regard to a standardized entity, e.g., like in our case, sentences or tokens. With this function you could, to give an example, calculate the average number of grammatical errors in learner texts that were rated B1 as compared to texts rated B2 (define your subcorpora first).

The procedure for displaying relative frequencies is analogue to the procedure mentioned above for absolute frequencies. Again, from the output on *<<Statistics >>* you can access all examples in the corpus in the *<<Advanced search>>* output by clicking on your search result.

(3)   error-free learner language

These measures depart from a positive perspective towards learner language in that the percentage of language that does **not** contain any of the annotated error-based features is displayed. You can calculate the percentage of error-free sentences or tokens with regard to the total number of sentences or tokens. On a more fine-grained level, you can also find out the percentage of language without any grammatical, morphological, or capitalization errors, for example.

(4)   complexity measures (German only)

For German, it was possible to include automatically calculated measures of morphological, lexical, and syntactical aspects of complexity in the statistics section. For Italian and Czech, unfortunately, the technical prerequisites were not given.

Complexity is an important aspect of (learner) language that has a close relationship to proficiency. In many studies, it has been shown that complexity is quite clearly distinguishable from accuracy and fluency (so-called 'CAF' studies, see bibliography for many references). The measures that MERLIN users have access to stem from research on L2 complexity and readability assessment. They were first implemented by Hancke (2013) and Hancke & Meurers (2013) with regard to MERLIN data *(<<documentation>>).*

5. Help

Graphic 37: Help

There are many documents and functionalities to help you find your way through the MERLIN interface. You can access the help section ( ) from anywhere on the interface.

5.1 User manual

This document, the user manual, is available on the <<help>> page on the MERLIN interface in German, Italian, English, and Czech and will be continuously updated.

5.2 Getting to know MERLIN: video introduction

You can watch two video registrations (each about 30 minutes) in which an overview of the MERLIN project is given (in English, by K. Wisniewski) . One introduction is directed towards language teachers, the second one is more appropriate for language testers or textbook authors. The presentations were registered during two workshops that the MERLIN team carried through in Linz in December 2014.

5.3 MERLIN interface navigation: The screencast tutorial

The screencast tutorial is available in English and German and guides you through the main functionalities of the MERLIN interface.

The tutorial has two blocks: a basic first part *(<<Define a subcorpus>>, <<Simple search>>)* enables you to get started by putting together and/or downloading texts / tasks you are interested in and search for words in them. In a second part, explanations on how to run an *<<Advanced search>>* on lemmas or annotations are given, and the *<<statistics>>* section is briefly introduced.

5.4 How to apply MERLIN: usage scenarios

In December 2014, the MERLIN team organized multiplier workshops in Linz, Austria, in which the MERLIN interface was introduced with the help of exemplary usage scenarios directed towards language teachers, testers, and trainers. These materials are freely available for download in the *<<Using MERLIN>>* section, as well, in Czech, Italian, and German. Please be aware of the fact that by the time the workshops were carried through, some MERLIN functionalities had not been implemented yet so that the scenarios might look slightly different from the current status of the MERLIN interface.

5.5 Understanding the <<Advanced search>> output

As mentioned above (chapter 4.3), the *<<Advanced search>>* uses the open source search and visualization architecture ANNIS. In the search output, in addition to the built-in features of ANNIS, the MERLIN team inserted some help functionalities to make it easier to understand.

Graphic 38: Help in output of exemplary <<Advanced search>>

47

By clicking on *<<about this search output>>,* an extra window opens so that you can continue to analyze the search output. All tiers of the Advanced search output grid are explained here (graphic 39).

Graphic 39: <<about this search output>> help function in <<Advanced search>> output

Another help option is to open a list with the *<<Abbreviations>>* used (graphic 40). The annotation tags with short definitions are displayed in an extra window.

Graphic 40: List of abbreviations used in the <<Advanced search>> output

5.6 Glossary

In the general glossary that you find in the help section ( ), terms used on the MERLIN interface are explained, many of them related to annotation.

5.7 Frequently asked questions

Here, questions by users are collected. You find information on what to use MERLIN for, on what to do with your search results, and on the *<<Advanced search>>* output, for example. This list will be continuously updated.

5.8 Contact us

Do not hesitate to contact the MERLIN team with any question or comment that arises (info@merlin-platform.eu). We are happy to help.

References

[ALTE 2001] = ALTE Working Group on the Code of Practice: *Principles of Good Practice for ALTE Examinations.* Revised Draft. http://www.testdaf.de/institut/pdf/ALTE/ALTE_good_practice.pdf, December 2014.

[Consiglio d'Europa 2004a] = Trim, J./North, B./Coste, D.: *Quadro comune europeo di riferimento per le lingue: apprendimento, insegnamento, valutazione*. La Nuova Italia: Oxford.- A cura del Consiglio d'Europa. [Council of Europe 1975] = Van Ek, J. A.: *The Threshold Level in a European unit/credit system for modern language learning by adults*. Strasbourg: Council of Europe.

[Council of Europe 1994a] = North, B.: *Scales of language proficiency: a survey of some existing systems.* Strasbourg: Council of Europe, CC-Lang (94) 24.

[Council of Europe 1994b [1981]] = Galli de' Paratesi, N.: *Livello Soglia per l'insegnamento dell'italiano come lingua straniera.* Strasbourg: Edizioni del Consiglio d'Europa.

[Council of Europe 1999 [1980]] = Baldegger, M./Müller, M./Schneider, G. (1999): *Kontaktschwelle Deutsch als Fremdsprache.* 4. Auflage. Berlin u.a.: Langenscheidt.- ed. by Council of Europe.

[Council of Europe 2001a] = Trim, J./North, B./Coste, D.: *Common European Framework of Reference for Languages: Learning, teaching, assessment*. -Edited by the Council of Europe. Online-Dokument: www.coe.int/lang, December 2014.

[Council of Europe 2001b] = Trim, J./North, B./Coste, D.: *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin u.a.: Langenscheidt.- Herausgegeben vom Europarat, Online-Dokument: http://www.goethe.de/z/50/commeuro/i7.htm, December 2014.

[Europarat 2004] = Takala, S./Kaftandjieva, F./Verhelst, N./Banerjee, J./Eckes, T./van der Schoot, F.: *Reference Supplement to the Preliminary Pilot Version of the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment*.- Edited by the Council of Europe, Online-Dokument: www.coe.int/lang, December 2014.

[Europarat 2009 [2003]] = North, B./Figueras, N./Takala, S./Van Avermaet, P./Verhelst, N.: *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. *Manual*. *Preliminary Pilot Version*.- Edited by the Council of Europe, Online-Dokument: www.coe.int/lang, December 2014.

Abel, A. / Wisniewski, K. / Nicolas, L. / Boyd, A. / Hana, J. / Meurers, D. (2014): A Trilingual Learner Corpus illustrating European Reference Levels. In: Ricognizioni – Rivista di Lingue, Letterature e Culture Moderne 2 (1), 111-126. (http://www.ojs.unito.it/index.php/ricognizioni).

Abel, A. / Glaznieks, A. / Nicolas, L. / Stemle, E. (2014): KoKo: an L1 Learner Corpus for German. In: Calzolari, N. / Choukri, K. / Declerck, Th. / Loftsson, H./ Maegaard, B. / Mariani, J. / Moreno, A. / Odijk, J. / Piperidis, St. (eds.): Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014), 26–31 May 2014. Reykjavik: European Language Resources Association (ELRA). 2414-2421. (http://www.lrec-conf.org/proceedings/lrec2014/index.html).

Abel, A. / Wisniewski, K. (2012): Sprechaktrealisierungen in der L2 und der GERS: Ein- und Aussichten für Sprachwissenschaft und Didaktik. In: Di Meola, Claudio / Hornung, Antonie / Rega, Lorenza (Hrsgg.): Perspektiven Vier. Akten der 4. Tagung Deutsche Sprachwissenschaft in Italien. Rom, 4.-6.Februar 2010. Frankfurt a.M. 311-325.

Abel, A. (2010): Sprachtests und soziale Implikationen. In: Deutsch als Fremdsprache, 4/2010. 202-209.

AERA/APA/NCME (1999): *Standards for educational and psychological testing.* Washington: AERA. Alderson, J.C. (1991): Bands and scores. In: Alderson, J.C./North, B. (eds.): *Language testing in the 1990s. London: British Council/Macmillan*, 71-86.

Aguado, K. (2004): Evaluation fremdsprachlicher Wortschatzkompetenz: Funktionen, Prinzipien, Charakteristika, Desiderate. In: Tschirner (Hrsg.) 231- 250.

Aijmer, K. (1996), *Conversational routines in English. Convention and creativity*. L*ondon*/NY: Longman.

Alderson, J. C./Figueras, N./Kuijper, H./Nold, G./Takala, S./Tardieu, C. (2006): Analysing Tests of Reading and Listening in Relation to the Common European Framework of Reference: The Experience of the Dutch CEFR Construct Project. In: *Language Assessment Quarterly* 3(1), 3-30.

Alderson, J.C. (2007): The CEFR and the need for more research. In: *The Modern Languagre Journal* 91, 658-662.

Al-Gahtani, S./Roever, C. (2012): Proficiency and Sequential Organization of L2 REQUESTs. In: Applied Linguistics 33/1, 42 –65.

Al-Jarf, R. Spelling error corpora in EFL. In the proceedings of the International Conference on Multi Development and Application of Language and Linguistics, National Cheng Kung University, May 5-16 2009, Tainan City, Taiwan, 2009.

Arnaud, P. J. L. (1984): The lexical richness of L2 written productionos and the validity of vocabulary tests: In: Culhane, T./Klein-Braley, C./Stevenson, D. K. (eds.): *Practice and Problems in Language*

Arnaud, P.J.L. (1984): The lexical richness of L2 written productions and the validity of vocabulary tests. In: Culhane, T./Klein-Braley, C./Stevenson, D.K. (eds.): *Practice and Problems in Language Testing.* Essex: Department of Language and Linguistics, University of Essex, 14-28.

Arras, U. (2010): Subjektive Theorien als Faktor bei der Beurteilung fremdsprachlicher Kompetenzen. In: Berndt, A./Kleppin, K. (eds.): *Sprachlehrforschung: Theorie und Empirie - Festschrift für Rüdiger Grotjahn*. Frankfurt: Lang, 169-179.

Bachman, L.F. (1990): *Fundamental Considerations in Language Testing*. New York: OUP.

Bachman, L.F. (2004): *Statistical analyses for language assessment*. Cambridge: CUP 2004.

Bachman, L.F./Palmer, A. (1996): *Language Testing in Practice*. New York: OUP.

Bachman, L.F./Palmer, A. (2010): *Language Testing in Practice. Developing Language Assessment and Justifying their Use in the Real World*. Oxford: OUP.

Bachmann, T. (2002): *Kohäsion und Kohärenz: Indikatoren für Schreibentwicklung: Zum Aufbau kohärenzstiftender Strukturen in instruktiven Texten von Kindern und Jugendlichen.* Innsbruck: Studienverlag.

Baker, P. (2010): *Sociolinguistics and Corpus Linguistics*. Edingburgh: EUP.

Bardovi-Harlig, K. & Bofman, T. (1989) Attainment of syntactic and morphological accuracy by advanced language learners. Studies in Second Language Acquisition, 11 (1), 17-34.

Bardovi-Harlig, K. (2009): Conventional Expressions as a Pragmalinguistic Resource: Recognition and Productions of Conventional Expressions in L2 Pragmatics. In: *Language Learning* 59 (4), 755-795.

Bardovi-Harlig, K. (2013): Developing L2 Pragmatics. In: Language Learning 63 (1): Suppl. 1, 66-86.

Barron, A. (2003), Acquisition in Interlanguage Pragmatics. Learning How to do Things with Words in a Study Abroad Context. Amsterdam/Philadelphia: Benjamins.

Bausch, K.-R./Christ, H./Königs, F.G./Krumm, H.-J. (eds.) (2003): *Der Gemeinsame Europäosche Referenzrahmen für Sprachen in der Diskussion. Arbeitspapiere der 15. Frühjarskonferenz zur Erforschung des Fremdsprachenunterrichts.* Tübingen: Narr.

Bayley, R. (2007): Second language acquisition: a variationist perspective. In: Bayley, R. /Lucas, C. (eds.) (2007): *Sociolinguistic Variation : Theories, Methods, and Applications.* Cambridge University Press, 133-144.

Bayley, R. / Regan, V. (eds) (2004). *The acquisition of sociolinguistic competence*. Special Issue of the *Journal of Sociolinguistics*, 8 (3).

Beebe, L. (1988). Five sociolinguistic approaches to Second Language Acquisition. In: L.Beebe (ed.), *Issues in Second Language Acquisition: Multiple Perspectives*. Cambridge, MA: Newbury House, pp. 43–75.

Bestgen, Y./Granger, S. (2011): Categorising spelling errors to assess L2 writing. In: *International Journal of Continuing Engineering Education and Life Long Learning,* 21 (2), 235-252.

Biber, D./Finegan, E. (eds.) (1994): *Sociolinguistic perspectives on register*. New York: OUP.

Blum-Kulka, S. (1987), Indirectness and politeness in requests: Same or different? *Journal of Pragmatics* 11, 1, 131-46.

Blum-Kulka, S. (1991), Interlanguage pragmatics: The case of requests. In Phillipson, R./Kellerman, E./Selinker, L./Sharwood Smith, M./Swain, M. (eds.) (1991), *Foreign/second language pedagogy research: A*

Blum-Kulka, S./House, J./Kasper, G. (eds.) (1989), Cross-cultural pragmatics: Requests and apologies. Norwood, NJ: Ablex.

Blum-Kulka, S./Olshtain, E. (1984), Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). Applied Linguistics 5, 3, 196-213.

Bond, T. G./Fox, C. M. (2007): Applying the Rasch model: Fundamental measurement in human sciences. Mahwah, NJ: Lawrence Erlbaum.

Botley, S. and Dillah, D. (2007) Investigating spelling errors in a Malaysian learner corpus. Malaysian Journal of ELT Research, Vol. 3, pp.74-93.

Bredel, U. (2010) (ed.) : Schriftsystem und Schrifterwerb: linguistisch – didaktisch – empirisch. Berlin: de Gruyter.

Bulté, B./Housen, A. (2012): Defining and operationalising L2 complexity. In: Housen, A./Kuiken, F./Vedder, I. (eds.): D*imensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Amsterdam: Benjamins, 21-46.

Burger, H. (2007): *Phraseologie. Eine Einführung am Beispiel des Deutschen*. (3. Aufl.).Berlin: Erich Schmidt Verlag.

Carlsen, C. (2010): Discourse connectives across CEFR levels: A corpus-based study. In: Bartning, I./Martin, M./Vedder, I. (eds.): *Communicative Proficiency and Linguistic Development: intersections between SLA and language testing research* (Eurosla). 191-210. purl.org/net/Carlsen-10.pdf.

Carlsen, C. (2010; Linking a learner corpus to the Common European Framework of Reference. Manuscript submitted for publication.

Carlsen, C. (ed.) 2013. *Norsk Profil. Det felles europeiske rammeverket spesifisert for norsk. Et første steg*. Oslo: Novus.

Casanave, C. (1994) Language development in students´journals. Journal of Second Language Writing. 3, 179-201.

Castro, C. D. (2004): Cohesion and the social construction of meaning in the essays of Filipino college students' writings in L2 English, in: Asia Pacific Education Review, 5, 215-225.

Chastain, K. (1990) Characteritics of graded and undergraded compositions. Modern Language Journal, 74, 10-14.

Chen, M., Zechner, K. (2011) Computing and Evaluating Syntactic Complexity Features for Automated Scoring of Spontaneous Non-Native Speech. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Portland, Oregon, June 19-24, 2011. ACL, pp. 722–731.

Chiang, St. (2003): The importance of cohesive conditions to perceptions of writing quality at the early stages of foreign language learning, in: System, Vol.31(4), 471-484.

Cho, Y. (2005) :Grammatik und Höflichkeit im Sprachvergleich. Direktive Handlungsspiele des Bittens, Aufforderns und Anweisens im Deutschen und Koreanischen. Tübinge.

Christ, O. (1994). A modular and flexible architecture for an integrated corpus query system. *arXiv preprint cmp-lg/9408005*.

Christ, O. (1994). A modular and flexible architecture for an integrated corpus query system. *arXiv preprint cmp-lg/9408005*.

Cook, V. (2005): Second Language writing systems. Clevedon: Multilingual Matters.

Cook, V. J. (1997). L2 users and English spelling. Journal of Multilingual and Multicultural Development, 18(6), 474-488.

Corder, S. P. (1993 [1973]): *Introducing Applied Linguistics*. Harmondsworth: Pelican. Dallapiazza, R.M./von Jan, E., Schönherr, T. (1998) (eds.): Tangram: *Deutsch als Fremdsprache. Kurs- und Arbeitsbuch 1 A*. Munich: Hueber.

Cornish, F. (2009): Inter-sentential anaphora and coherence relations in discourse: a perfect match, in: Language Science 31 (2009), 572-592.

Coulmas, F. (Hrsg.) (1986): Direct and Indirect Speech., Berlin, New York, Amsterdam.

Crossley, S./Salsbury, T./McNamara, D. (2009): Measuring L2 Lexical Growth Using Hypernymic Relationships. In: *Language Learning* 59 (2) 307-334.

Crossley, Sc.A./McNamara, D.S. (2011): Shared features of L2 writing: Intergroup homogeneity and text classification, in: Journal of Second Language Writing 20 (4) 271-285.

Dale, E. (1965): Vocabulary measurement: Techniques and major findings. In: *Elementary English* 42, 895-901.

Daller, H./Milton, J./Treffers-Daller, J. (eds.) (2007): *Modelling and Assessing Vocabulary Knowledge*. Cambridge: CUP.

Daller, H./van Hou, R./Treffers-Daller, J. (2003): Lexical richness in spontaneous speech of bilinguals. In: *Applied Linguistics* 24, 197-222.

Dellapiazza, R.M./von Jan, E., Schönherr, T. (1998) (Hrsgg.): Tan*gram: Deutsch als Fremdsprache. Kurs- und Arbeitsbuch* 1 A. Munich: Hueber.

Dewaele, J.-M. (2004): Indiviual differences in the use of colloquial vocabulary. The effects of sociobiographical and psychological factors. In: Bogaards, P./Laufer, L. (eds.): Vocabulary in a secons language. Amsterdam: John Bejamins, 127-154.

Dewaele, J.-M. (2004): The acquisition of sociolinguistic competence in French as a foreign language: An overview. In: *French Language Studies* 14, 301–319.

Díaz-Negrillo, A./Fernández-Domínguez, J. (2006): Error-coding systems for learner corpora. In: *RESLA* 19, 83-102.

Durrant, P./Schmitt, N. (2009): To what extent do native and non-native writers make use of collocations? In: *International Review of Applied Linguistics in Language Teaching (IRAL)*, 47 (2) 157-177.

Durrant, P./Schmitt, N. (2009): To what extent do native and non-native writers make use of collocations? In: *International Review of Applied Linguistics in Language Teaching (IRAL)*, 47 (2) 157-177.

Eckes, T. (2008): Rater types in writing performance assessments: A classification approach to rater variability. In: *Language Testing 25* (2) 155-185.

Eckes, T. (2009): *Reference Supplement to the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Section H: Many-Facet Rasch Measurement*. (http://www.coe.int/t/dg4/linguistic/manuel1_en.asp, December 2014.)

Eisenberg, P. (2007): Sprachliches Wissen im *Wörterbuch der Zweifelsfälle*. über die Rekonstruktion einer Gebrauchsnorm. In: *Aptum. Zeitschrift für Sprachkritik und Sprachkultur* 3/2007: 209-228.

Ellis, R. (1994): *The study of Second Language Acquisition*. Oxford: Oxford University Press.

Fabricius-Hansen, C. (2005): Elusive connectives. A case study on the explicitness dimension of discourse coherence. Linguistics, 43, 17-48.

Fatemi, M. A. (2008) The relationship between writing competence, language proficiency and grammatical errors in the writing of Iranian tefl sophomores. Doctoral thesis. Universiti Sains Malaysia.

Fender, M. Spelling knowledge and reading development: Insights from Arab ESL learners. Reading in a Foreign Language, 2008, Volume 20, No. 1, pp. 19–42.

Feng, L. (2010): Automatic readability assessment. PhD thesis, City University of New York (CUNY), http://gradworks.umi.com/3426751.pdf (February 2015).

Foster, P./Tavakoli, P. (2009): Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. In: *Language Learning* 59 (4) 866-896.

Fulcher, G. (2004): Deluded by Artifices? The Common European Framework and Harmonization. In: *Language Assessment Quarterly* 1 (4), 253-266.

Fulcher, G./Davidson, F. (2007): *Language Testing and Assessment.* London/New York: Routledge. Gould, S.J. (1996): *The mismeasure of man*. London: Penguin.

Gernsbacher, M. A./Givón, T. (Eds.) (1995): Coherence in Spontaneous Text. Benjamins, Amsterdam.

Glaznieks, A. / Nicolas, L. / Stemle, E. / Abel, A. / Lyding, Verena (2014): Establishing a Standardised Procedure for Building Learner Corpora – a Response to Demands and Suggestions of Users. In: *Apples - Journal of Applied Language Studie*s 8 (3), 2014 (http://apples.jyu.fi/issue/view/15)

Graesser, A. C./Millis, K. K./Zwaan, R. (1997): Discourse comprehension, in: Annual Review of Psychology 48, 163-189.

Granger, S. (2002): A Bird's-eye view of learner corpus research. In: Granger S,/Hung, J./ Petch-Tyson, St (eds.): *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teachin*g. Amsterdam: John Benjamins, 3-33.

Granger, S. (2003): Error-tagged learner corpora and CALL: a promising synergy. In: *CALICO Journal* 20 (3). Special issues on error analysis and error correction in computer-assisted language learning, 465-480.

Granger, S. (2008): Learner corpora. In: Lüdeling, A. / Kytö, M. (eds.): *Corpus linguistics: an international handbook* (Handbooks of linguistics and communication science; 29.1_ 29.2). Berlin - New York: de Gruyter. 259-275.

Granger, S., Bestgen, Y (2011) Categorizing spelling errors to assess L2 writing. International Journal of Continuing Engineering Education and Life Long Learning, 21, 2-3, 2011, 235 - 252.

Guiraud, P. (1954): *Les caractères statistiques du vocabulaire.* Paris: Presse Universitaires de France.

Gyllstad, H./Granfeldt, J./Bernardini, P./Källkvist, P. (2014): Linguistic correlates to communicative proficiency levels oft he CEFR. The case opf syntactic complexity on written L2 English, L3 French and L4 Italian. In: *EUROSLA Yearbook* 14, 1-30.

Halliday, M. A. K. /Hasan, R. (1989): *Language, context and text: a social semiotic perspective.* Oxford: Oxford University. Press.

Halliday, M. A. K./Hasan, R. (1976): Cohesion in English. London, Longman.

Hana, J./Rosen, A./ Štindlová, B./Štěpánek, J. (2014): Building a learner corpus. In: Language Resources and Evaluation 8, SE

Hana, J./Rosen, A./Skodová, S. Stindlová, B.: Error-tagged learner corpus of Czech. In: *Proceedings of the Fourth Linguistic Annotation Workshop,* ACL 2010, Uppsala, Sweden 2010, 11-19.

Hancke J./Meurers D./Vajjala S. (2012): Readability Classification for German using lexical, syntactic, and morphological features*.* In: *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, 1063-1080.

Hancke, J. *Automatic Prediction of CEFR Proficiency Levels Based on Linguistic Features of Learner Language*. Master's thesis, Universität Tübingen, April 2013.

Hancke, J./Meurers, D./Vajjala, D. (2012): Readability classification for German using lexical, syntactic, and morphological features. In: Proceedings of the 24th International Conference on Computational Linguistics (COLING), 1063-1080, Mumbay, India.

Hasil, J./Hájková, E./Hasilová, H. (2007): *Brána jazyka českého otevřená*. Prague: Karolinum.

Hasko, V. (2013): Capturing the Dynamics of Second Language Development via Learner corpus research: a very long engagement. In: *The Modern Language Journal* 97, S1, 1-10.

Hattingh, K. (2005) The syntactic development of grade 12 ESL learners. Dissertation, North-West University, Potchefstroom Campus.

Hawkey, R./Barker, F. (2004): Developing a Common Scale for the Assessment of Writing. In: Assessing Writing 9, 122-159.

Hawkins, J. A./Filipovíc, L. (2012): *Criterial features in L2 English: Specifying the reference levels of the Common European Framework*. Cambridge: CUP.

Held, G. (1995): Verbale Höflichkeit. Studien zur linguistischen Theorienbildung und empirische Untersuchung zum Sprachverhalten französischer und italienischer Jugendlicher in Bitt-und Dankessituationen. Tübingen: Narr.

Holmes/Brown 1976: Developing sociolinguistic competence in a second language. *Tesol Quarterly* 10/4.

Housen, A., Kuiken, F. (2009) Complexity, Accuracy, and Fluency in Second Language Acquisition. Applied Linguistics, 30(4): 461-473

Housen, A./Kuiken, F. (2009): Complexity, Accuracy, and Fluency in Second Language Acquisition. In: *Applied Linguistics* 30 (4) 461-473.

Housen, A./Kuiken, F./Vedder, I. (eds.) (2012): *Dimensions of L2 Performance and Proficiency Investigating Complexity, Accuracy and Fluency in SLA*. Amsterdam: John Benjamins.

Hovermale, DJ., Martin, S. Developing an Annotation Scheme for ELL Spelling Errors. Proceedings of MCLC-5 (Midwest Computational Linguistics Colloquium) East Lansing, Michigan, USA, May 10-11, 2008.

Hudson, Tom/Detmer, Emily/Brown, J.D. (1992): A framework for testing cross-cultural pragmatics. Hawaii: University of Hawai'I at Manoa.

Hudson, Tom/Detmer, Emily/Brown, J.D. (1995): *Developing prototypic measures of cross-cultural pragmatics.* Hawaii: University of Hawai'I at Manoa.

Hulstijn, J. H. (2007): The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. In: *The Modern Language Journa*l 91, 663-667.

Hulstijn, J. H./Alderson, C./Schoonen, R. (2010): Developmental stages in second-language acquisition and levels of second-language proficiency: Are there links between them? In: Bartning, I./Martin, M./Vedder, I. (eds.): *Communicative Proficiency and Linguistic dvelopment: intersections between SLA and language testing research*. Eurosla Monograph Series. (http://eurosla.org/monographs/EM01/EM01home.html*)*

Hymes, D. (1974): *Foundations in Sociolinguistics*. Philadelphia: University of Pennsylvania Press.

Ishikawa, S. (1995) Objective Measurement of Low-Proficiency EFL Narrative Writing. Journal of Second Language Writing, 4: 51 - 70.

Jarvis, S. (2002): Short texts, best-fitting curves and new measures of lexical diversity. In: *Language Testing* 19 (1) 57-84.

Jelínek, T., Barbora Štindlová, Alexandr Rosen, Jirka Hana (2012). Combining Manual and Automatic Annotation of a Learner Corpus. Text, Speech and Dialogue Lecture Notes in Computer Science Volume 7499, pp 127-134http://ufal.mff.cuni.cz/~hana/bib.html#rosen-etal-2013-czesl-lre, December 2014.

Johns, T. (1988): Whence and whither classroom concordancing? In: Bongaarts, T./de Haan, P./Lobbe, S./Wekker, H. (eds.): Computer Applications in Language Learning. Dordrecht: Foris, 9-33.

Johns, T. (1997): Contexts: The Background, Development and Trialling of a Concordance-based CALL Program. In: Wichmann, Anne/Fligelstone, Steven/McEnery, Tony/Knowles, Gerry (eds.) (1997), *Teaching and Language Corpora.* London: Longman, 100-115.

Kaczmarek, C. M. (1980) Scoring and Rating 'essay tasks'. (& Oller, J.W. and Perkins, K. &. Research in Language Testing. Rowley, Massachusetts: Newbury House.)

Koch, P./Oesterreicher, W. (2011²) : *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch.* Berlin u.a.: de Gruyter.

Kroll, B. (1990) What does time buy? ESL student performance on home versus class compositions. In B. Kroll, ed. Second language writing: Research insight for the classroom. Cambridge: CUP, pp. 140 – 154.

Kuiken, F./Vedder, I./Gilabert, R. (2010): Communicative Adequacy and Linguistic complexity in L2 writing. EUROSLA Monographs Series 1, 81-100.

Larsen-Freeman, D. (2009): Adjusting Expectations: The Study of Complexity, Accuracy, and Fluency in Second Language Acquisition. In: *Applied Linguistics* 30 (4) 579-589.

Laufer, B. (1995): Beyond 2,000. A measure of productive lexicon in a second language. In: Eubank, L./Selinker, L./Sharwood Smith, M. (eds.): The *Current State of Interlanguage.* Amsterdam/Philadelphia: John Benjamins, 265-272.

Laufer, B./Nation, P. (1995): Vocabulary size and use: lexical richness in L3 written production. In: *Applied Linguistics* 16, 307-322.

Linnarud, M. (1986): *Lexis in composition: A performance analysis of Swedish learners' written English.* Malmö: CWK Gleerup.

Little, D. (2007): The Common European Framework of Reference for Languages: Perspectives on the Making of Supranational Languages Education Policiy. In: *The Modern Language Journal* 91, 645-655.

Lorenzo-Dus, N. (2007): The best of both worlds? Combined methodological approaches to the assessment of vocabulary in oral proficiency interviews. In: Daller/Milton/Treffers-Daller (eds.) 220-233.

Louwerse, M.M./Graesser, A. C. (2004). Coherence in discourse, in: Strazny, P. (Ed.): Encyclopedia of linguistics. Chicago: Fitzroy Dearborn.

Lu, X. (2009). Automatic measurement of syntactic complexity in child language acquisition. International Journal of Corpus Linguistics 14, 3–28(26). URL http://www.ingentaconnect.com/content/jbp/ijcl/2009/00000014/00000001/art00002.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. International Journal of Corpus Linguistics, 15(4):474-496.

Lu, X. (2011): A corpus-based evaluation of syntactic complexity measures as indices of College-level ESL writers' language development. In: *TESOL Quarterly* 45 (1) 36-62.

Lu, X. (2012): The relationship of lexical richness to the quality of ESL learners' oral narratives. In: *The Modern Language Journal* , 190-208.

Lüdeling, A. (2008): Mehrdeutigkeiten und Kategorisierung: Probleme bei der Annotation von Lernerkorpora. In: Walter, M./Grommes, P. (eds.): *Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitsprachenerwerbsforschung.* Tübingen: Niemeyer, 119-140.

Lüdeling, A./Walter, M./Kroymann, E./Adolphs, P. (2005): Multi-level Error Annotation in Learner Corpora. In: Hunston, S./Danielsson, P. (eds.): *Proceedings from the Corpus Linguistics Conference Series* (Corpus Linguistics 2005, Birmingham, 1415 July 2005). (http://www.corpus.bham.ac.uk/PCLC).

Lyster, R. (1994). The effect of functional-analytic teaching on aspects of French immersion students' sociolinguistic competence. *Applied Linguistics*, 15 (3) 263–287.

Malvern, D./Richards, B./Chipere, N./Durán, P. (2008²): *Lexical Diversity and Language Development. Quantification and Assessment.* New York: Palgrave Macmillan.

Matsuda, P. K. (1997): Contrastive rhetoric in context: A dynamic model of L2 writing, in: Journal of Second Language Writing, 6, 45-60.

McCarthy, P.M./Jarvis, S. (2007): vocd: A theoretical and empirical evaluation. In: *Language Testing* 24 (4) 459-488.

McNamara, D. S./Kintsch, E./Butler Songer, N./Kintsch, W. (1996): Are Good Texts Always Better? Interactions of Text Coherence, Background Knowledge, and Levels of Understanding in Learning from Text, in: Cognition and Instruction, Vol. 14, No. 1 (1996), 1-43.

McNamara, D.S./Kintsch, W. (1996): Learning from text: Effects of prior knowledge and text coherence, in: Discourse Processes, 22, 247-287.

McNamara, D.S./Louwerse, M.M. /Graesser, A.C. (unpublished): Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension. Grant proposal.

Mellor, A. (2011): Essay Length, Lexical Diversity and Automatic Essay Scoring. In: *Memoirs of the Osaka Institute of Technology*, Series B Vol. 55, No. 2 (2011), 1-14.

Ménard, N. (1983): *Mesure de la richesse lexicale. Théorie et vérifications expérimentales. Études stylométriques et sociolinguistiques*. Genf/Paris: Slatkine-Champion.

Meurers, D. (2012): Natural Language Processing and Language Learning. *Encyclopedia of Applied Linguistics*. Blackwell. purl.org/dm/papers/meurers-11.html.

Mezzadri, M. (2000): *Rete! Book 1.* Perugia: Guerra Edizioni.

Mougeon, R./Dewaele, J.-M. (2004): Preface. In: IRAL 42 (4) (Special Issue: Variation in the interlanguage of advanced second language learners.], 295-301.

Mougeon, Raymond/Nadasdi, Terry/Rehner, Katherine (2010): *The Sociolinguistic Competence of Immersion Students.*

Müller, Ch./Strube M. (2006): Multi-Level Annotation of Linguistic Data with MMAX2. In: S. Braun, K. Kohn,J. Mukherjee (Eds.): *Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods.* Frankfurt: Peter Lang, 197-214.

Nassaji, H. (2003). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. The Modern Language Journal, 87, 261-276.

Nation, P. (2001): *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, P. (2007): Fundamental issues in modelling and assessing vocabulary knowledge. In: Daller, H./ Milton, J./Treffers-Daller, J. (eds.): *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press.

Nesselhauf, N. (2005): *Collocations in a Learner Corpus*. Amsterdam: John Benjamins. North, B. (2000): *The Development of a Common Framework Scale of Language Proficiency.* Oxford: Peter Lang.

North, B. (2000): *The Development of a Common Framework Scale of Language Proficiency.* Oxford: Peter Lang.

Nuzzo, E. (2007): Imparare a fare cose con le parole. Richieste, proteste, scuse in italiano lingua seconda. Perugia: Guerra.

Nuzzo, E. (2009): "Buongiorno, ho bisogno dell"informazione per andara a barcellona": uno studio longitudinale sulle richieste di informazioni e suggerimenti in italiano L2. In: Linguistica e Filologia 28, 83 –109.

O'Loughlin, K. (1995): Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. In: *Language Testing* 12 (2) 217-237.

Okada, T. A Corpus Analysis of Spelling Errors Made by Japanese EFL Writers.

Ortega, L. (2003) Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. Applied Linguistics, 24 (4), 492-518.

Ortega, L. (2012): Interlanguage complexity: A construct in search of theoretical renewal. In: Szmrecsanyi, B./Kortmann, B. (Eds): *Linguistic complexity in interlanguage varieties, L2 varieties, and contact languages*. Berlin: Walter de Gruyter.

Ott, N. (2009). Information Retrieval for Language Learning: An Exploration of Text Difficulty Measures. Master's thesis, University of Tubingen, Seminar fur Sprachwissenschaft, Tübingen, Germany. URL http://drni.de/zap/ma-thesis.

Paquot, M./Granger, S. (2012): Formulaic language in Learner Corpora. In: *Annual Review of Applied Linguistics* 32, 130-149.

Perfetti, C. A., Rieben, L. & Fayol, M. (eds) (1997). *Learning to Spell: Research, Theory, and Practice across Languages*. Mahwah, NJ: Lawrence-Erlbaum Associates.

Petersen, S.E./Ostendorf, M. (2009): A machine learning approach to reading level assessment. In: *Computer Speech and Language* (23) 86-106.

Polio, C. (1997). Measures of linguistic accuracy in second language writing research. Language Learning, 47, 101-143.

Pollitt, A./Murray, N.L. (1996): What raters really pay attention to. In: Milanovic, M./Saville, N. (eds.): *Performance testing, cognition and assessment; Selected papers from the 15th Language Testing Research Colloquium.* Cambridge: Cambrudge University Press, 74-91.

Preston, D. (2000). Three kinds of sociolinguistics and SLA: A psycholinguistic perspective. In B. Swierzbin, F. Morris, M. E. Anderson, C. E. Klee and E. Tarone (eds), *Social and cognitive factors in second language acquisition*. Somerville: Cascadilla Press, pp. 3–30.

Read, J. (2000): *Assessing vocabular*y. Cambridge: Cambridge University Press.

Read, J. (2007): Second language vocabulary assessment: current practice and new directions. In: *International Journal of English Studies* 7 (2) 105-125.

Read, J./Chapelle, C. (2001): A framework for second language vocabulary assessment. In: *Language Testing* 18, 1-32.

Read, J./Nation, P. (2004): Measurement of formulaic sequences. In: Schmitt, N. (ed.): *Formulaic sequences: Acquisition, processing and use.* Amsterdam: John Benjamins, 23-35.

Regan, V. (1995). The acquisition of sociolinguistic native speech norms. In: B. Freed (ed.), *Second language acquisition in a study abroad context*. Amsterdam, Philadelphia: Benjamins, pp. 245–267. (vorgemerkt)

Regan, Vera/Howard, Martin/Leme, Isabelle (2009): *The Acquisition of Sociolinguistic Competence in a Study Abroad Context.*

Reznicek, M./Lüdeling, A./Hirschmann, H. (in print): Competing Target Hypotheses in the Falko Corpus. A Flexible Multi-Layer Corpus Architecture. In: Díaz-Negrillo, A./Ballier, N./Thompson, P. (eds.): *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam: John Benjamins (Series Studies in Corpus Linguistics).

Reznicek, M./Lüdeling, A./Krummes, C./Schwantuschke, F./Walter, M./Schmidt, K./Hirschmann, H./Andreas,T. (2012): *Das Falko-Handbuch. Korpusaufbau und Annotatione*n. Version 2.01. HU Berlin (http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko/Falko-Handbuch_Korpusaufbau%20und%20Annotationen_v2.01)

Rimrott, A., Heift, T. (2008). Evaluating automatic detection of misspellings in German. Language Learning & Technology.

Römer, U. (2010): Using general and specialized corpora in English language teaching: past, present and future. In: Campoy-Cubillo, M. et al. (eds.): Corpus-based approaches to English Language Teaching. London: Continuum, 18-38.

Römer, Ute. 2008. 7. Corpora and language teaching. In: Lüdeling, Anke & Merja Kytö (eds.). *Corpus Linguistics. An International Handbook (volume 1)*. [HSK series] Berlin: Mouton de Gruyter. 112-130.

Römer. U. (2006): Pedagogical applications of corpora: some reflections on the current scope and a wish list for future developments. In: Zeitschrift für ANglistik und Amerikanistik 54 (2) 121-134.

Rose, K.R. (2000), An exploratory cross-sectional study of interlanguage pragmatic development. *Studies in Second Language Acquisition 22,* 1, 27-67.

Rosen, A., Jirka Hana, Barbora Štindlová, and Anna Feldman (2013): Evaluating and automating the annotation of a learner corpus. *Language Resources and Evaluation*, pages 1-28, April 2013. http://ufal.mff.cuni.cz/~hana/bib.html#rosen-etal-2013-czesl-lre, December 2014.

Sanders, T./Pander Maat, H. (2006): Cohesion and coherence: Linguistic approaches, in: Brown, K., et al. (Eds.), Encyclopedia of Language and Linguistics. Elsevier, London.

Sassoon, R. (1995). *The Acquisition of a Second Writing System*. Oxford: Intellect.

Schmitt, N. (ed.) (2004): *Formulaic Sequences: Acquisition, Processing, and Use*. Amsterdam: John Benjamins Press.

Schmitt, N./Carter, N. (2004): Formulaic sequences in action: An Introduction. In: Schmitt, N. (ed.): *Formulaic sequences: Acquisition, processing and use.* Amsterdam: John Benjamins, 1-21.

Schneider, G./North, B. (2000): *Fremdsprachen können - was heißt das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit.* Nationales Forschungsprogramm 33, Wirksamkeit unserer Bildungssysteme. Chur, Zürich: Rüegger.

Schneider, J. G. (2013): Sprachliche ‚Fehler' aus sprachwissenschaftlicher Sicht. In: *Sprachreport* 1-2/2013, 30-37.

Skehan, P. (2009): Modelling Second Language Performance: Integrating complexity, accuracy, fluency and lexis. In: *Applied Linguistics* 30 (4) 510-532.

Škodová, S.,Barbora Štindlová, Jirka Hana and Alexandr Rosen (2011). Víceúrovňová anotace českého žákovského korpusu. In: V. Petkevič and A. Rosen (Eds.), *Korpusová lingvistika Praha 2011: 3 - Gramatika a značkování korpusů,* 16. pp 208-225. Studie z korpusové lingvistiky. Nakladatelství Lidové noviny:Praha.

Spinelli, B./Parizzi, F. (ed.) (2010): *Profilo della lingua italiana*. Firenze: La Nuova Italia. Stede, M. (2007): Korpusgestützte Textanalyse. Grundzüge der Ebenen-orientierten Textlinguistik. Tübingen: Narr.

Spooren, W./Sanders, T. (2008): The acquisition order of coherence relations: On cognitive complexity in discourse, in: Journal of Pragmatics 40 (2008), 2003-2026.

Štindlová B., S. Škodová, A. Rosen and J. Hana (2012). Annotating foreign learners' Czech In: Studies in Formal Slavic Linguistics. Contributions from Formal Description of Slavic Languages 8.5, Peter Lang GmbH, Frankfurt am Main, Germany, pp. 205-219.

Taguchi, N. (2012): *Context, individual differences and pragmatic competence*. Bristol: Multilingual Matters.

Tarone, E. (2007): Sociolinguistic approaches to second language acquisition research, 1997-2007. In: *Modern Language Journal* 91*,* 837-848.

Trosborg, A. (1995), Interlanguage pragmatics: Requests, complaints, apologies. Berlin: Mouton de Gruyter.

UIMA Framework: http://uima.apache.org (July 2013)

Ure, J.N. (1971): Lexical density and register differentiation. In: Perren, G.E./Trim, J.L.M. (eds.): *Applications of Linguistics: Selected papers of the Second International Congress of Applied Linguistics.* Cambridge: CUP, 443-452.

Vajjala, S./Meurers, D. (2012): On improving the accuracy of readability classification using insights from second language acquisition. In: Joel Tetreault, Jill Burstein, and Claudial Leacock, editors, Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications (BEA7) at NAACL-HLT. Montreal, Canada, June 2012, 163-173. http://aclweb.org/anthology/W12-2019.pdf (February 2015).

van Compernolle, Remi A. /Williams, Lawrence (2012): Teaching, Learning, and Developing L2 French Sociolinguistic Competence: A Sociocultural Perspective *Applied Linguistics (*2012*) 33(2): 184-205 first published online January 24, 2012.*

Vaughan, C. (1991): Holistic assessment: What goes on in the rater's mind? In: Hamp-Lyons L. (ed.): *Assessing Second Language Writing in Academic Contexts.* Norwood: Ablex, 111.125.

Vedder, I.(2007): Competenza pragmatic e complessità sintattica in italiano L2: l'uso dei modificatori nelle richieste. In: Linguistica e Filologia 25, 99-123.

Vivanco, V. (2005): The absence of connectives and the maintenance of coherence in publicity texts, in: Journal of Pragmatics 37 (2005), 1233-1249.

Warga, M. and U. Scholmberger (2007). The acquisition of French apologetic behaviour in a study abroad context. Intercultural Pragmatics, 4, 221-251.

Wesche, M./Paribakht, T.S. (1996): Assessing second language vocabulary knowledge depth versus breadth. In: *The Canadian Modern Language Review* 53, 13-40.

Wisniewski, K. (2010): Bewertervariabilität im Umgang mit GeRS-Skalen. Ein- und Aussichten aus einem Sprachtestprojekt. In: *Deutsch als Fremdsprache* 3, 143-150.

Wisniewski, K. (2012): Lexikalische Kompetenzen in der Fremdsprache testen: Ein Modellierungsansatz. In: In: Abel, A. / Vettori, C. / Wisniewski, K. (eds.): Gli studenti altoatesini e la seconda lingua: indagine linguistica e psicosociale. / Die Südtiroler SchülerInnen und die Zweitsprache: eine linguistische und sozialpsychologische Untersuchung. Volume 2 – Band 2. Bolzano - Bozen: Eurac. ., 24-49. (http://www.eurac.edu/en/research/publications/PublicationDetails.aspx?pubId=0100156&type=Q)

Wisniewski, K. (2013): The empirical validity of the CEFR fluency scale: the A2 level description. In: Galaczi, E.D./Weir, C.J. (eds.): *Exploring Language Frameworks: Proceedings of the ALTE Krakow Conference*. Cambridge: Cambridge University Press, 253-272. Studies in Language Testing.

Wisniewski, K. (2014): *Die Validität der Skalen des Gemeinsamen europäischen Referenzrahmens für Sprachen. Eine empirische Untersuchung der Flüssigkeits- und Wortschatzskalen des GeRS am Beispiel des Italienischen und des Deutschen*. Frankfurt: Peter Lang. Language Testing and Evaluation Series, 33.

Wisniewski, K. / Abel, A. (2012): Die Sprachkompetenzerhebung: Theorie, Methoden, Qualitätssicherung. In: Abel, A. / Vettori, C. / Wisniewski, K. (eds.): *Gli studenti altoatesini e la seconda lingua: indagine linguistica e psicosociale. / Die Südtiroler SchülerInnen und die Zweitsprache: eine linguistische und sozialpsychologische Untersuchung.* Volume 1 – Band 1. Bolzano - Bozen: Eurac. 13-64 (http://www.eurac.edu/en/research/publications/PublicationDetails.aspx?pubId=0100156&type=Q)

Wisniewski, K./Schöne, K./Nicolas, L./Vettori, C./ Boyd, A./Meurers, D./ Abel, A./Hana, J. (2013): MERLIN: An online trilingual learner corpus empirically grounding the European Reference Levels in authentic learner data. In: *ICT for Language Learning, Conference Proceedings 2013*. Libreriauniversitaria.it Edizioni. (http://conference.pixel-online.net/ICT4LL2013/common/download/Paper_pdf/322-CEF03-FP-Wisniewski-ICT2013.pdf)

Wolfe-Quinteiro, K., Inagaki, S., Kim, H.-Y. (1998) Second language development in writing: measures of fluency, accuracy, & complexity. Hawai'i: University of Hawai'I at Manoa.

Wray, A. (2002): *Formulaic Language and the Lexicon.* Cambridge: CUP.

Yang, W./Sun, Y. (2012): The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. In: *Linguistics and Education*, 23 (1), 31-48. Wray, A. (2002): *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Young, R. (1999). Sociolinguistic approaches to SLA. *Annual Review of Applied Linguistics*, 19: 105–32.

Yu, Ming Chung (2005): Sociolinguistic Competence in the Complimenting Act of Native Chinese and American English Speakers: A Mirror of Cultural Value. In: *Language and Speech* 48 (1) 91-119.

Yu, Ming-chung (2011): Learning how to read situations and know what is the right thing to say or do in an L2: A study of socio-cultural competence and language transfer. *Journal of Pragmatics*, 2011, Vol.43(4), psp.1127-1147.

Zeldes, A./Ritz J./Lüdeling A. et al. (2009): *Annis: A search tool for multi-layer annotated corpora. In Proceedings of Corpus Linguistics*, July 20-23. Liverpool. (http://ucrel.lancs.ac.uk/publications/cl2009/).

Zipser, F./Romary, L./al. (2010). A model oriented approach to the mapping of annotation formats using standards. In: *Workshop on Language Resource and Language Technology Standards, LREC 2010*.

Zuskin, Robin D. (1992): Assessing L2 Sociolinguistic Competence: In Search of Support from Pragmatic Theories. In: *Pragmatics and Language Learning* 4.