



**Internet tools for learning level-appropriate text  
selection in German as a Foreign Language?**

Uwe Fricke (Shanghai)

ISSN 1470 – 9570

## Internet tools for learning level-appropriate text selection in German as a Foreign Language?<sup>1</sup>

Uwe Fricke (Shanghai)

The aim of this study is to examine whether text analysis tools can be helpful in the selection of factual texts for learners of German as a Foreign Language in Asia. Text analysis tools make it possible to determine the level of difficulty of texts according to so-called readability formulae. A specially compiled collection of 50 essay texts were ranked by using these tools according to their level of difficulty. Out of these 50 texts six were chosen for a pair-wise comparison. University students in Thailand and China, majoring in German (n=47), assessed the texts in terms of their relative difficulty. Teachers of those students (n=10) were asked to do the same. The broad agreement of the rankings determined by the tools and by the participants shows that the use of internet tools can facilitate the selection of suitable texts.

### 1. Introduction

Teachers of German as a Foreign Language and especially of German Studies face the question of which texts are appropriate for the learning level of their learner group. This is especially true when the selection should rather be based on the topic (e.g. tourism in the home country, culture of the west) than aiming at general facilitation of the four skills within language learning. The process of the hermeneutic understanding of literary texts will not be of interest here either, although there may not be a clear boundary between factual and literary or fictional texts.

Demands on the selection of factual texts are usually that they should be current and interesting. In addition, the demand for the use of authentic texts has played an important role for quite some time (Thonhauser 2010: 1034). This is also shown by a look at the “*Handbuch Fremdsprachendidaktik*“ (2013<sup>2</sup>), in which the keyword ‘authenticity’ is one of the most frequently used terms in the register. Finding up-to-date and authentic texts has become easier in the age of the internet, so that a series of texts on a wide variety of topics

---

<sup>1</sup> The research was carried out at Thammasat University (Bangkok, Thailand) and presented in German at the International Conference of the Indonesian Association of Germanists on September 29, 2018 in Yogyakarta, Indonesia. I want to thank Jonathan Majumder for his proofreading of the draft.

can be searched for in a relatively short amount of time. But then there is still the criterion that is at issue in this study, namely the difficulty of texts.

## 2. Research Problem

The problem of selecting a learning level appropriate reading text has occupied research for some time. In the series “Research into Practice”, the educational scientist Sharon Murphy (2013: 1) briefly summed up the most important insights into the interplay of factors that make up the textual difficulty for the reader (here talking about children):

- “knowledge about the reader’s characteristics and the reading task
- knowledge about the surface features of a text
- knowledge about the deeper features of texts and the modalities represented in the text.“

Leaving aside the contributions of traditional hermeneutics, research roughly divides into those approaches that are primarily concerned with surface features (quantitative approach), and those which, in critical dismissal, focus on deeper features of texts as well as the relationship between text and reader (qualitative approach). The former is also referred to as readability research, the latter as text intelligibility research (Groeben & Christmann 1989).

The surface features were the focus of interest in the middle of the 20th century. The best-known research of this kind resulted in the so-called Flesch formula, which was developed by the Austrian Rudolf Flesch for the determination of the difficulty of English-language texts (1948). It is still used alongside alternative formulae today. For the German language there are modified versions of the Flesch formula.

The so-called Flesch Reading Ease for German is based on a modified formula by Toni Amstad:  $180 - (\text{number of words divided by number of sentences}) - (58.5 \text{ times } [\text{number of syllables divided by number of words}])$ . An alternative readability index is the *Läsbarhetsindex* (Swedish, and briefly known as LIX) developed by the Swedish scholar Carl-Hugo Björnsson. This is calculated by the  $[\text{total number of words divided by the number of sentences}] \text{ plus } [\text{long words with more than 6 letters multiplied by 100 and divided by the total number of words}]$  or in short: the sum of the average sentence length of a text and the percentage of long words. These readability formulae are also processed by those text analysis tools that are used here in the first part of this study as an instrument for measuring textual difficulty.

In the English-speaking discourse the approach to capturing readability by means of the surface features was continuously present. In the German-speaking discourse of the 1990s, however, this approach was considered unsuitable for the guidance of teaching practice (Nebe 1990) and therefore not to be pursued further (Groeben & Christmann 1998: 167).

It is only recently, in the German context, that surface features of texts have become a more popular research topic (e.g. Oelke et al. 2012, Hancke et al. 2012). This has been driven by information technology and the automation of processes with the associated possibilities for in-depth text analysis by incorporating further text features.

My reason for returning to the topic can be seen in a dilemma of text intelligibility research (*Textverständlichkeitsforschung*). In the dilemma of the conflicting goals of precision and practicality, this study tends to be orientated towards practicality. Despite the negative results of Ursula Nebe (1990) with regard to practical handling, it seems necessary to re-examine whether readability formulae processed by tools can provide practical help with text selection. It is because of the complexity of the subject that the alternative construct which is used in the study of text intelligibility could not yet be transformed into “manageable criteria” for teaching practice, as demanded by Nebe (1990).

Three aspects should be mentioned in this context. First, hardly any studies on textual difficulty have been conducted with foreign language learners, even in the anglophone sphere. This was already mentioned by Nebe (1990), but was still seen as a challenge in 2016 (Xia et al.). Secondly, the method of pairing used in this study, which is often used in other fields of research, has hardly been used in the given context: According to De Clercq et al. (2012), apart from their own study on Dutch there is only one more by Tanaka-Ishii et al. (2010) with corpora in English and Japanese. Thirdly, it could be added that a holistic assessment of the difficulty of the text by those who are really the subject of the investigations, the learning readers, is hard to find. I am only aware of the study “Using the Crowd for readability Prediction” (De Clercq et al. 2012). In many cases, expert estimates have been used as a reference criterion, since the results obtained by means of readability formulae have been criticized. Groeben, who himself follows a theoretical-deductive approach, outlines as typical of an empirical-inductive way that “texts of different levels of difficulty from different areas are estimated by experts with the help of an impression differential with regard to relevant text features” (1989: 169). After this rather general localization and legitimization of the research problem a concrete investigation design is described.

### 3. Research Aim

The aim of the investigation is to clarify whether text analysis tools can be of help in the selection of texts appropriate for learning German in Asia. An explanation should be provided by a comparison regarding the evaluation of textual difficulties a) through text analysis tools on the internet and b) a subjective, rather holistic assessment of students of German studies in Asia.

The tools process a few surface features of the texts and give results with values between 0 and 100. While we have no insight into the black box of the students' subjective assessments, presumably they also rely on surface features, including their knowledge of the vocabulary used. The students skimmed over two texts simultaneously (switching between them) in a paired comparison of about 45 seconds, and then they decided which text seemed more difficult. The results collected from the students' paired comparisons made it possible to form a ranking. The index values for the texts, which were given by the tools, also resulted in a ranking. A comparison of the rankings by means of text analysis tools on the one hand and by the paired comparison of the students on the other provides information about the suitability of the text analysis tools for the learning level-appropriate text selection. It should be noted here that the students are only concerned with an ad hoc forecast of how they judge the textual difficulty. Within such a forecast, the interest of readers in a topic, their motivation to read about it and/or their prior background knowledge should be among other important factors in estimating text difficulty. And so the motivation and interest of the learner is surely an important factor for an appropriate text selection by the teacher. However, no statement can be made about how difficult it is for the reader to actually read and understand the text following the ad hoc estimation. And within this experiment no data will be generated concerning the multifaceted construct of motivation (see an overview by Gambrell 2011:6f.), which is (due to Konradi et al. 2013) lacking clarity in the overall research discourse. In the pairwise comparison, a test person could be comparatively more motivated to read one of the two texts, which may give him the impression that this text is easier to read, especially if the difference in text complexity between the two texts is rather small.

## 4. Research Design and Method

### 4.1 Study Participants

A total of 47 students majoring in German (plus 10 of their teachers) participated in the experiment. 38 subjects studied (resp. 10 worked) at two Bangkok-based universities. In addition, 9 participants from two universities in China participated in the experiment. The students, in their second, third or fourth year of study, were aged between 19 and 23 years old, and the number of female participants was much higher than that of male students. The participants were acquired via the teachers of the departments. In the case of the Chinese students, a contact person among the Chinese students was responsible for conducting the experiment. The course of the experiment was largely standardized by an approximately 5 minute presentation explaining the experiment and after that performing the experiment with a timed presentation of about 20 minutes.

### 4.2 Experimental Design and Survey Instruments

An *analysis by text analysis tools* on the internet is easy to perform and is based on a readability formula. The text is copied into a text window and within a few seconds, it produces a readability index value. Five free instruments for German-language texts (see listing at the end) were used and their values – all on the same scale between 0 and 100 – were averaged for each text. The results of the tools differed slightly, as it is not always clear whether the basic formula has been modified by incorporating further text features. Three of the tools work with a Flesch formula for German and the results from these tools are not significantly different from each other (standard deviations between 1 and 4 points). In one case the Flesch Reading Ease for English is used, in another case the *Läsbarhetsindex* (LIX). Including these tools brings about higher standard deviations (between 6 and 10 points), nevertheless the instruments were offered for the measurement of texts in German. Since this study is not about a particular tool or formula or the search for the best tool, the five values have been averaged and serve as a basis for comparison with the assessments of the (potential) readers of the texts, the students of German. However, using the average of the three tools based on German or the average value of all five tools does not make a difference for the ranking of the texts, but it certainly could be of importance when discussing a calibration of tools.

For the assessment of the texts by the students a *pairwise comparison* was used. The pairwise comparison is a method that is particularly used in medical, social science or consumer research and dates back to a pioneer of psychometry, Louis Leon Thurstone. Among other things, it allows the researcher to bring a set of objects (in this case texts) step by step into a preference ranking in complex decision-making processes. The various criteria that underlie a voting decision can also be used without the subjects' awareness, because there are only two alternatives. Finally, a preference ranking can be determined from the sum of the individual decisions, which may well be inconsistent. A disadvantage of this method is that, realistically, only a limited number of objects can be included, since with each additional object, the number of necessary comparisons increases rapidly. Each object is compared with every other object. But a high number of subjects could quickly experience a fatigue effect, which would impair the quality of the results. For the pair comparison, a timed slide presentation was used. The participants saw two texts in comparison and each individual ticked a prepared form for each comparison, confirming which text was more difficult. Each slide, which was shown for about 45 seconds, contained in the title the question to be answered: "Which text do you find harder?" The arrangement of the pairs was made according to the algorithm proposed for pair comparisons by the psychologist RT Ross (1934), the principle of equilibrium – each pair element appears as often as possible on the left and right sides – and the principle of maximum distance – the distance between the first appearance of a pair element and its repeated appearance should be as large as possible (Cloete et al. 1988).

### 4.3 Text Material

The text selection took place in stages. At the beginning, about 35 factual texts on different topics from different sources were collected for the experiment. In addition to texts from daily and weekly newspapers (*Die Welt, Die Zeit, sternNEO, Süddeutsche Zeitung*, etc.), texts from internet material for children and adolescents (*Planet Wissen, Fluter*, etc.) have been included in order to cover a wider range of difficulty. Restricted authentic texts from internet sites for learners of German, including especially *Deutsche Welle*, were added. *Deutsche Welle* is Germany's international broadcaster financed by federal tax resources and also provides access to the German language (<https://www.dw.com/en>). For some of these texts, learning level information due to the *Common European Framework of Reference* (CEFR) was given. This provides an opportunity for a rather marginal

comparison of “expert” assessments with the results of the automatic readability-index given by the tools.

The number of texts increased to a total of 50 in order to achieve the fullest possible range of difficulty levels and sufficient choice in each difficulty area. In this second phase of text collection, texts were examined to close gaps in some value ranges and a few texts with high difficulty were supplemented from scientific sources.

The index values range from 0 to 100, with a low index value representing a high degree of difficulty of the text and a high value corresponding to a rather low degree of difficulty. This is the opposite of the LIX, but it is easy to handle the reciprocal. Index values above 90 and below 30 were never reached and appear to be virtually non-existent. The average index values of the five analysis tools were used for a ranking that can be compared to that based on the experiment with the participants.

Since in a pair comparison realistically only six text objects can be compared, which means at least 15 pair comparisons were necessary, among the 50 texts six texts had to be selected for the pair comparison. Originally, the goal was to include texts with as even a spacing between their index values as possible. In the selection of the six texts for the experiment, however, mistakenly the index values for the respective complete text were still received, whereas in the experiment only the beginning of the text was used. The texts were cut for pragmatic reasons to the same length of about 190 words (+/- 10), on the one hand to prevent the evaluation of the students being influenced by the text length and on the other hand to avoid effects of fatigue among the participants. And to exclude the side effects of layout, the texts were formatted identically. Apart from the text, there was only the title, the source was not specified. The fact that the results of the tools, namely the index values could still change when shortening was not considered. For example, the text may have higher or lower index values if the average length of sentences in the shortened version changed to a certain degree. But given time pressure, this could not be changed. But the resulting small deviations – the text might get index values – hardly affected the final result. However, the index values of the chosen six selected texts were no longer as originally intended in approximately equal distance from each other. Another goal, which should be realized with the text selection, was, to represent a thematic spectrum because the interests of the reader and also an existing knowledge of the vocabulary must be seen in this experiment as an essential conditioning factor in reading comprehension. In the selection for the pair comparison, the following texts were used according to the above criteria (in

brackets the rounded mean for the readability index, after which the source is mentioned for the reader of this article):

- A. Toleranz in der Zeit der Reformation** (41) Projekt “DenkWege zu Luther”  
(*English*: Tolerance in the Time of the Reformation)
- B. Auch Schimpansen können schmollen** (46) Deutsche Welle (C1/C2)  
(*English*: Even chimpanzees can pout)
- C. Erfinder: Leonardo da Vinci, das Universalgenie** (47) Planet Wissen  
(*English*: Inventor: Leonardo da Vinci, the universal genius)
- D. Punktesystem für brave Bürger** (54) SZ-Magazin  
(*English*: Points system for good citizens)
- E. Die Welt des Fahrrads** (59) Deutsche Welle (C1/C2)  
(*English*: The World of the Bicycle)
- F. Matterhorn** (69) K(inder)lexikon  
(*English*: Matterhorn)

Particularly simple or difficult texts, i.e. texts from the periphery of the value spectrum, were not included. This increases the claim for suitability of the tools for text selection because the smaller the distances between the texts, the more demanding the need for accuracy is. The texts in the pair comparison are 10 points or fewer apart. If one wanted to assign the index values to the levels of the *Common European Framework of Reference* (CEFR), this could be done – as here suggested – on a trial basis shown in the figure (Fig. 1) in steps of 10 index points (in the experiment from B1 to C2) because the border area is rather theoretical and without practical value. Before working with it in other contexts, this classification needs further discussion and would have to be confirmed or adjusted using a larger data set.

CEFR levels	Flesch index
	100
A1	90
A2	80
B1	70
B2	60
C1	50
C2	40
	30
	20
	10
	0

**Fig. 1: allocation of CEFR levels and Flesch index**  
[in the border area without practical value]

Fig. 1: CEFR levels in relation to Flesch index

## 5. Results

This section presents the results of three comparisons: first the focal question will be answered by the comparison of the ranking by students with those by internet tools, secondly the ranking of students and teachers will be compared, and the third comparison concerns the practical usage of the readability index versus recommendations of publishers of learner texts.

Is there a correlation between the rankings provided by the internet tools and those of the students? The results to the central question of this study can be shown in a rank order table (Tab. 1), here with average index values for all 5 tools and for the 3 tools with less standard deviation.

Which text is more difficult?	Internet-Tools		Students (n=47)
	Index Value Ø of 5 (Ø of 3)	Rank	Rank
A. Toleranz in der Zeit der Reformation	<b>41,2</b> (42,7)	<b>1</b>	<b>1</b>
B. Auch Schimpansen können schmollen	<b>46,4</b> (50,7)	<b>2</b>	<b>2</b>
C. Erfinder: Leonardo da Vinci, das Universalgenie	<b>47,4</b> (51,3)	<b>3*</b>	<b>4*</b>
D. Punktesystem für brave Bürger	<b>54,0</b> (58,3)	<b>4*</b>	<b>3*</b>
E. Die Welt des Fahrrads	<b>59,0</b> (63,0)	<b>5</b>	<b>5</b>
F. Matterhorn	<b>69,2</b> (75,0)	<b>6</b>	<b>6</b>

Positive Rank Correlation: Kendall's Tau = 0.867,  $\alpha = 0.01$

As already mentioned, the problem of a lack of equidistance amongst the index values is evidenced by texts B and C (gray background), which are only one point apart so that it was to be expected that the test persons would not recognize this difference.

However, a difference in the comparison of the ranking places is shown (surprisingly) elsewhere, namely in the change in the rankings of the texts C and D, i.e. in the medium range of difficulty of the text selection (each with \*). Both the simpler and the two more difficult texts show a match in the rank assignment.

In a comparison of the teachers and students, the teachers ranked the texts in the same order as the students (Tab. 2 on the following page). The result, however, shows a more

differentiated gradation among the texts compared to the result among the students. This can be seen in the percentage figures enclosed in the additional column (rounded to 0.5 percent each), which show the proportion of points awarded in the paired comparison. (The paired as harder classified text got each one point.) Here it is clear that the distance among the three classified as difficult texts in the total group of students is very low (between 21 and 22 percent, each with gray background). The verdict of the students taken as a group is quite uncertain in this range, where, according to them, the three texts are almost at the same level.

In general, the students who took part in the experiment had a limited age range and were part of a small group at the respective university. If one compares these small groups (6 in total), it turns out that in almost all of these six small groups the texts F “Matterhorn” and E “Die Welt des Fahrrads” (The World of the Bicycle ) (with only one exception and text E on rank 4) received the ranks 6 and 5. But the ranks for the two most difficult texts go to the texts A, B, C and D in very different constellations when comparing the six groups.

Tab. 2: Ranking Comparison: Tools, Students & Teachers

<i>Which text is more difficult?</i>	<b>Internet-Tools</b>	<b>Students</b> (n=47)		<b>Teachers</b> (n=10)	
	<b>Rank</b>	<b>Rank</b>	<b>%</b>	<b>Rank</b>	<b>%</b>
A. Toleranz in der Zeit der Reformation	<b>1</b>	<b>1</b>	<b>22</b>	<b>1</b>	<b>30</b>
B. Auch Schimpansen können schmollen	<b>2</b>	<b>2</b>	<b>21.5</b>	<b>2</b>	<b>20.5</b>
C. Erfinder: Leonardo da Vinci, das Universalgenie	<b>3</b>	<b>4</b>	<b>17.5</b>	<b>4</b>	<b>14.5</b>
D. Punktesystem für brave Bürger	<b>4</b>	<b>3</b>	<b>21</b>	<b>3</b>	<b>18</b>
E. Die Welt des Fahrrads	<b>5</b>	<b>5</b>	<b>13.5</b>	<b>5</b>	<b>12.5</b>
F. Matterhorn	<b>6</b>	<b>6</b>	<b>4</b>	<b>6</b>	<b>4</b>

Positive Rank Correlation: Kendall's Tau = 0.867,  $\alpha = 0.01$

The text D “Punktesystem für brave Bürger” (Points system for good citizens), which deals with China, is considered more difficult by the Thai students (as well by their teachers) than indicated by the values given by the internet ranking tools. Only the Chinese students rank the text in the same way as the internet tools. This may be interpreted as an indication

that the assessment of the difficulty may also be about thematic factors. We can assume that the Chinese students were more familiar with or more interested in the topic of their country than the students in Thailand.

A comparison between the index values of the analysis tools and the level recommendations corresponding to the *Common European Framework of Reference for Languages* (CEFR) for those included texts that originate from internet pages for learners of German brings another insight. Among the 50 texts of the collection were 18 texts with such a level indicator, 14 of them from Deutsche Welle, the German broadcaster, which offers reliable news coverage and access to the German language (others from *e-learning Alpen-Adria*, *GermanNet*, *Klett*, *Vitamin.de*). Among these texts, there were also some less accurate assignments, so ten texts were assigned to two levels and one even three levels. Nevertheless, these classifications are used here exploratively as expert ratings, even if no information is available on how these classifications came about. To indicate this, the term expert is inserted here in quotation marks. A request to this effect to *Deutsche Welle* remained unanswered.

If one sketches experimentally the allocation of levels by the publishers, authors or presumably experts as outlined above (Fig. 1, above) and identifies the specified levels in such an index value (with multiple assignment then interpreted as an intermediate value), the result is the following matrix (Fig. 2, below).

Six out of the 18 texts fall within the vicinity range of plus/minus ten points of the index value, and twelve do not fall within this range. (Following the suggestion mentioned above, ten points equal the difference between two levels.) The information provided by the “experts” of the learner texts are therefore less consistent with the index values of the internet tools, which in turn correspond quite well with the assessment by students and teachers.

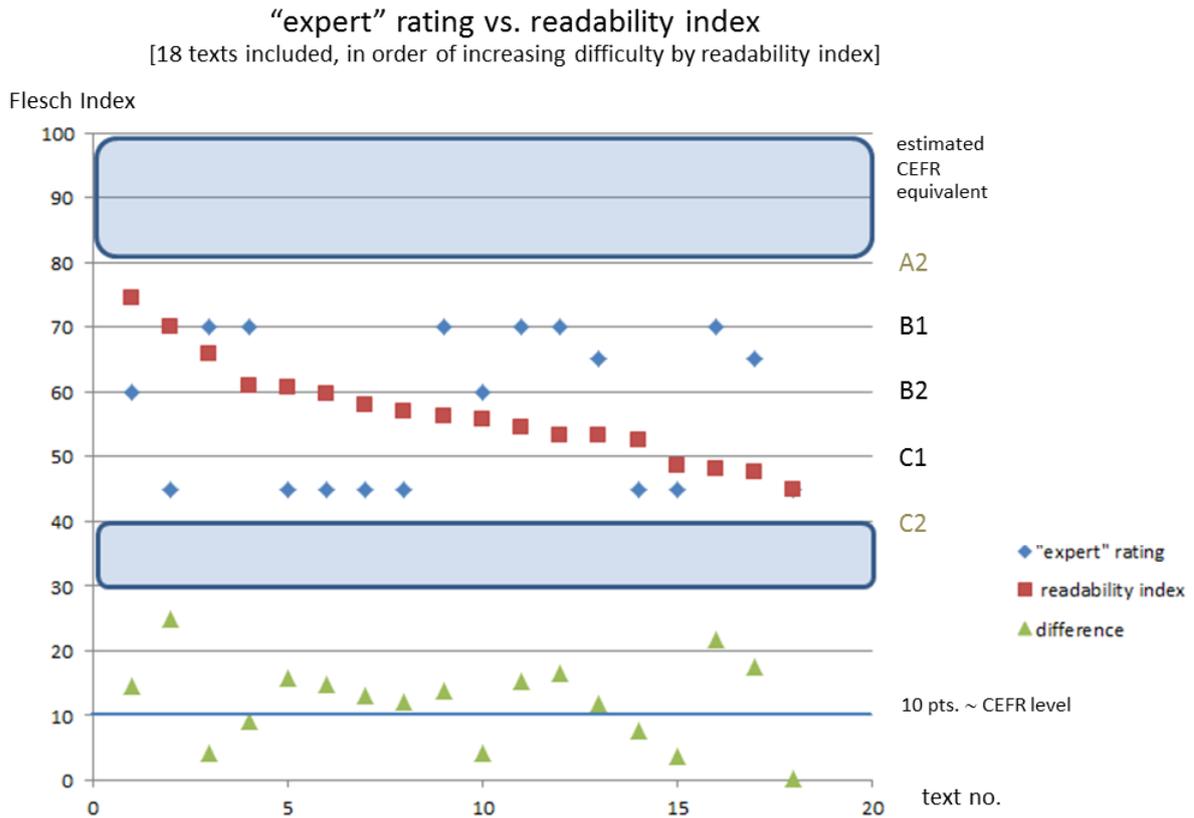


Fig. 2: “expert” rating (suggestions for texts offered to learners of German) versus readability index

## 6. Discussion & Conclusions

The aim of this study was to check the suitability of text analysis tools on the internet for text selection in the study of German. The experiment, in which students gave an ad hoc assessment of the textual difficulty in paired comparisons, showed that there is a rank correlation between the students’ assessment and the index values of the internet tools. This connection and the ability among the students to differentiate between the levels of difficulty is more evident in the simpler texts and decreases or even disappears as the difficulty of the texts increases. The explanation for this may be that the students have not yet reached the level of the more difficult texts, so it is much harder for them to assess differences. This probably requires further investigation, possibly including a preliminary test to determine the level of the students. There is also evidence of the influence of prior knowledge and interest in the topic, as the differing assessment of the Chinese students in the text on China may indicate. However, capturing thematic influences in an experiment is likely to be difficult because, in addition to texts used in the curriculum, individual dispositions and interests also play a role here.

The experiment with the teachers – who can be considered advanced “learners” – shows that the rank correlation is also present for the assessment of the more difficult texts. But here the number of subjects involved in the investigation may have been too small.

The experiment shows the hypothesis that text analysis tools represent a suitable aid in the learning level-appropriate text selection for German students cannot be refuted. The suitability of text analysis tools in the sense of a prognosis in the selection of text therefore needs to be examined further - also in view of the above-mentioned desiderata.

If we compare the information on textual difficulty given on some internet pages for learners of German with the orientation given by the index values of the analysis tools, the results presented here speak for the preference of the aid given by text analysis tools.

Outlook: The answer to the question of whether special analysis tools are needed for foreign language learners must remain open for the time being. The existing tools, which do not focus on foreign language learning, however, already provide sufficiently good results, the results within a comparison of tools indicate that using a special formula for German leads to better results, i.e., significantly smaller standard deviations between the results for the index values of the tools.

Of course, one can think of the inclusion and processing of information on the frequency of use of words in order to include aspects of the vocabulary of learners, as has already been discussed in the literature and is now also technically easy to implement. The German publisher *Klett* offers a tool named *Language Level Evaluator* for text analysis that among other characteristics of the text indicates the difficulty level of the words for learners corresponding to CEFR. Whether this leads to other index values for the textual difficulty has yet to be proven and as well as whether this leads to a better predictive power than the formula based tools. The result of this experiment has shown that the predictive power of the tools used seems to be already quite acceptable, at least as far as the ad hoc judgement and the rather motivational aspects of learners are concerned. The first impression of the text difficulty might be important for the attitude and motivation of a learner when starting to read a text. But the method used, based on ad hoc judgements, does not allow us to conclude more than that. The clearly more complex question of how difficult the actual process of working on a text is has, of course, not yet been answered.

But given some further confirmation within future investigations – by replication with a variation of texts and more participants – possibly a rather simple ranking tool could be

designed, which would include measurements of word frequency classes for generating a difficulty score in order to provide more reliable CEFR indication when it comes to the task of a selection of factual texts for learners of German at an intermediate level. We should not lose sight of the aspect of practicality, being aware that more advanced computational linguistics is possible. Teachers using such tools should be able to understand the procedures and therefore to judge the results themselves.

Changing the starting point of this study, the results can also provide information for the text selection of didactically long-term planned learning materials, i.e. course book texts. Thinking of the three options, index values based on formulae and processed by tools, expert judgements and learner judgements, we could also think of starting with learner judgements instead of the expert judgements or – as suggested here – with tools. Those who can and want to offer a huge text corpus for learners, usually textbook publishers, could make use – not only of tools but – of “using the crowd for readability prediction” as suggested by the title of De Clercq et al. As we saw, the learners’ judgements might differ from expert judgements. Furthermore, up to the date of the experiment there seemed to be no research attempting to measure the progression of text difficulties in textbooks for learners of German. Since the readership is large in this case, the effort and investment in such an enterprise, which could be done as an online experiment, could be worthwhile. Since the experiment, some time has passed and the authors of the Language Level Evaluator now also indicate to provide an "overall score for the difficulty of the text" (L-Pub 2020).

## References

- Cloete, W.G.; Cloete, I.; Gadow, Klaus von (1988) An algorithm for presenting pairs in optimum orders. In: *EDV in Medizin und Biologie*, 19. Jg. H. 2-3, 75-77. URL: <https://ul.qucosa.de/api/qucosa%3A12639/attachment/ATT-0/>.
- De Clercq, Orphee; Hoste, Veronique; Desmet, Bart; Van Oosten, Philip; De Cock, Martine; Macken, Lieve (2014) Using the crowd for readability prediction. In: *Natural Language Engineering*, 20 (3), 293-325. DOI: <https://doi.org/10.1017/S1351324912000344>.
- Gambrell, Linda B. (2011): Motivation in the School Reading Curriculum. In: *Journal of Reading Education*, 37 (1), 5-14. URL: <https://rdlg579.files.wordpress.com/2015/06/motivation-in-the-school-reading-curriculum-gambrell-copy.pdf>.
- Groeben Norbert; Christmann, Ursula (1989) Textoptimierung unter Verständlichkeitsperspektive. In: Antos, Gerd; Krings, Hans P. (eds.): *Textproduktion. Ein interdisziplinärer Forschungsüberblick*. Tübingen (=Konzepte der Sprach- und Literaturwissenschaft 48) 165-196.

- Hancke, Julia; Vajjala, Sowmya; Meurers, Detmar (eds.) (2012) Readability Classification for German using lexical, syntactic, and morphological features. In: *Association for Computational Linguistics: 24th International Conference on Computational Linguistics*. Bombay (17 pages without page numbers) URL: <http://www.sfs.uni-tuebingen.de/~dm/papers/Hancke.Vajjala.Meurers-12.pdf>.
- Konradi, Kristin; Jang, Bong Gee; Mckenna, Michael C. (2014) Motivation terminology in Reading Research: A Conceptual Review. In: *Educational Psychology Review*, 26 (1), 127-164. URL: [https://www.researchgate.net/publication/263491242\\_-\\_Motivation\\_Terminology\\_in\\_Reading\\_Research\\_A\\_Conceptual\\_Review](https://www.researchgate.net/publication/263491242_-_Motivation_Terminology_in_Reading_Research_A_Conceptual_Review).
- Oelke, Daniela; Spretke, David; Stoffel, Andreas; Keim, Daniel A. (2012) Visual Readability Analysis: How to Make Your Writings Easier to Read. In: *IEEE Transactions on Visualization and Computer Graphics*, 18. Jg. H. 5, 662-674. URL: <http://www.cs.tufts.edu/comp/250VA/papers/readability-oelke.pdf>.
- L-Pub (2020): Language Level Evaluator URL: <https://l-pub.com/2020/12/14/language-level-evaluator/>.
- Murphy, Sharon (2013) Assessing Text Difficulty for Students. In: *What Works? Research into Practice*, No. 44, Jan. 1-4. URL: [http://www.edu.gov.on.ca/eng/literacynumeracy/inspire/research/ww\\_atds.pdf](http://www.edu.gov.on.ca/eng/literacynumeracy/inspire/research/ww_atds.pdf).
- Nebe, Ursula (1990) Ist Textschwierigkeit messbar? In: *Deutsch als Fremdsprache*, 27. Jg. H. 6, 350-356.
- Richter, Tobias; Schnotz, Wolfgang (2018) Textverstehen. In: Rost, Detlef H.; Sparfeldt, Jörn R.; Buch, Susanne R. (eds.) *Handwörterbuch Pädagogische Psychologie*. Weinheim 826-837. URL: [https://www.psychologie.uni-wuerzburg.de/fileadmin/06020400/Textverstehen\\_Richter\\_Schotz\\_2018\\_preprint.pdf](https://www.psychologie.uni-wuerzburg.de/fileadmin/06020400/Textverstehen_Richter_Schotz_2018_preprint.pdf).
- Tanaka-Ishii, Kumiko; Tezuka, Satoshi, Terada, Hiroshi (2010) Sorting texts by readability. In: *Computational Linguistics*. Vol. 36 No. 2, 203-227. URL: <https://www.aclweb.org/anthology/J10-2002.pdf>.
- Thonhauser, Ingo (2010) Textarbeit. In: Krumm, Hans-Jürgen; Fandrych, Christian; Hufeisen, Britta; Riemer Claudia (eds.): *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch*. Berlin, New York, 1033-1039.
- Xia, Menglin; Kochmar, Ekaterina; Briscoe, Ted (2016) Text Readability Assessment for Second Language Learners. In: *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*. San Diego, California 12–22. URL: <https://www.aclweb.org/anthology/W16-0502.pdf>.

### **Texts (Excerpts) in the pair comparison:**

1. Erlinghagen, M. (23.08.2017): Die Welt des Fahrrads. In: Deutsche Welle, Alltagsdeutsch. URL: <https://www.dw.com/de/die-welt-des-fahrrads/1-40151105>.
2. Hacke, A. (15.12.2016): Punktesystem für brave Bürger. In: SZ-Magazin 49/2016. URL: <https://sz-magazin.sueddeutsche.de/das-beste-aus-aller-welt/punktesystem-fuer-brave-buerger-83121>.
3. Höck, D.; Passin, C. (2017): Reformation und Toleranz. Anregungen zum Philosophieren mit Jugendlichen aus dem Projekt „DenkWege zu Luther“ URL: [https://www.denkwege-zu-luther.de/papers/DenkWege\\_zu\\_Luther\\_Reformation\\_und\\_Toleranz\\_web.pdf](https://www.denkwege-zu-luther.de/papers/DenkWege_zu_Luther_Reformation_und_Toleranz_web.pdf).

4. K(inder)lexikon: Matterhorn. URL: <https://klexikon.zum.de/wiki/Matterhorn>.
5. Planet Wissen: Erfinder: Leonardo da Vinci – das Universalgenie. URL: <https://www.planet-wissen.de/technik/erfindungen/erfinder/pwieleonardodavinci-dasuniversalgenie100.html>.
6. Rößler, N.; Warken, B. (01.05.2018): Auch Schimpansen können schmollen. In: Deutsche Welle, Alltagsdeutsch. URL: <https://www.dw.com/de/auch-schimpansek%C3%B6nnen-schmollen/1-41597230>.

### **Text analysis tools (in parentheses the formula used):**

Wortliga GmbH (München, Deutschland): *Wortliga Textanalyse Tool*. (Flesch reading Ease according to Toni Amstad) URL: <https://wortliga.de/textanalyse/>.

Pierre D. (Leverkusen, Deutschland): *Textanalyse-Tool.de*. (Lesix, Flesch Reading Ease and other criteria) URL: <http://www.textanalyse-tool.de/>.

Bachmann, Ch. (Frauenfeld, Schweiz): *Leichtlesbar.ch*. (Original Flesch Reading Ease) URL: <http://www.leichtlesbar.ch/html/>.

Bookshouse Ltd. (Zypern): *Schreiblabor Textanalyse*. (diverse, shown separately: Original Flesch Reading Ease. German Formula of Flesch Reading Ease according to ToniAmstad et.al.) URL: <http://www.schreiblabor.com/textanalyse/>.

Klett derdieDaF-Portal: *LLE-Language Level Evaluator – der Niveau-Check für Ihre Texte* URL: <https://www.derdiedaf.com/lle-der-niveau-check-fuer-ihre-texte/c-191>.

Übersetzungsbüro Supertext (Zürich, Schweiz): *Supertext-LIX-Rechner*. (LIX according to Björnsson) URL: <https://www.supertext.ch/tools/lix>.

All internet links were last accessed on Aug 30, 2018

### **Keywords**

DaF, Germanic studies, learning level, readability formula, pair comparison, ranking, subjective reader judgment, text analysis tools, text selection, textual difficulty

### **Author Biography**

Uwe Fricke studied social sciences and German language at Leibniz University Hannover, he holds a diploma in social sciences and is currently working as a lecturer in German at Shanghai Dianji University in China. Experience as a lecturer for German as a foreign language in the Czech Republic, China and Thailand, as well as teaching and research in the social and educational sciences. Email: ufri13@web.de