RESEARCH ARTICLE

LATIC—A linguistic analyzer for text and item characteristics

Nadine Cruz Neri*, Florian Klückmann, Jan Retelsdorf

Faculty of Education, Educational Psychology, University of Hamburg, Hamburg, Germany

* nadine.cruz.neri@uni-hamburg.de

Abstract

Analyzing texts and items regarding their linguistic features might be important for researchers to investigate the effects of the linguistic presentation as well as for practitioners to estimate the readability of a text or an item. The Linguistic Analyzer for Text and Item Characteristics (LATIC) is a software that enables users to analyze texts and items more efficiently. LATIC offers a multitude of features at three different reading levels and can be used for texts and items in four different languages: English, French, German, and Spanish. It is open source, free to use and designed to be user-friendly. In this study, we investigated LATIC’s performance: LATIC achieves highly accurate results, while being extremely time saving compared to human raters. While developing LATIC, the respective features are tested continuously to ensure a high accuracy of results in the future.

Introduction

In the last few decades, many tools have been developed to analyze text and item characteristics regarding their linguistic and syntactical complexity. To name but a few, the tools TAACO (Tool for the Automatic Analysis of Cohesion) [1] and ReaderBench [2] are available for English texts and items. However, tools to date all have their own strengths, but might also have a few shortcomings regarding their features and accessibility for researchers and practitioners. For instance, some tools only allow the analysis of texts and items in one language (e.g., ARTE [Automatic Readability Tool for English] [3]) or fail to report accuracy measures of the implemented features (e.g., RATTE [Regensburg Analysis Tool for Texts] [4]).

This article introduces a new tool that tries to overcome some of these shortcomings: A Linguistic Analyzer for Text and Item Characteristics (LATIC). LATIC enables users to analyze texts and items automatically at three reading levels: the word, sentence, and text level. Using the Stanford CoreNLP [5], LATIC annotates parts of speech (e.g., nouns, verbs) and counts them. Additionally, LATIC can calculate different traditional readability indices (e.g., Flesch Reading Ease [6]; Läserbarhetsindex [LIX] [7]) and other objective measures (e.g., average sentence length, word count). To date, LATIC is available for English, French, German, and Spanish and can be used on most operating systems without a registration or even an installation being needed.

In this article, we aim to introduce LATIC and highlight its features. First, we illustrate why it is important to analyze linguistic text and item characteristics and how LATIC can help users in doing so. Therefore, we show how LATIC can be used and which features LATIC
access or privileges that other authors would not have. Second, we worked with lists containing the most frequent words in English and German to examine the accuracy of the syllable count. For testing the German version, we worked with a list provided by the Leipzig Corpora Collection (Goldhahn et al., 2012), which can be downloaded under https://wortschatz.uni-leipzig.de/de/download/german. Again, we cannot share this list due to copyright reasons, but we confirm that other authors could access the data in the same way we did. We did not have any special access or privileges that other authors would not have. For the English version, we created a list ourselves. https://en-syllables.latic.software/. Third, we worked with ratings of human coders. These are openly available on https://rating-results.latic.software/.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

(version 1.2.2) provides to date. Second, we investigate LATIC’s performance in terms of accuracy and time savings (compared to human raters). Finally, in the discussion we elaborate on advantages as well as limitations of LATIC comparing it to other tools.

The importance of analyzing linguistic text and item characteristics

It is widely known that reading processes depend on reader characteristics, such as reading skills [8–10], but also on text and item characteristics [11, 12]. Regarding the latter, especially the linguistic complexity has gained a lot of attention in the past decades [13, 14]. In order to ensure that readers are able to comprehend written texts and items, it is essential to consider how linguistically complex these texts and items are. Many researchers and educators rely on traditional readability indices [15], such as the Flesch Reading Ease [6] and the SMOG (Simple Measure of Gobbledygook) [16] to estimate the complexity of a text or an item. However, it is argued that these readability indices might not be appropriate predictors of linguistic complexity since they are usually based on a few variables only, such as word and sentence length [17]. In fact, it is known that text and item characteristics that are not implemented into these readability indices can also generate difficulties in comprehension. The number of prepositional phrases, for instance, hinders comprehension and is associated with lower performance of students [18] and adults [12]. Hence, it is important not to rely on traditional readability indices only [17], but to analyze texts and items in more depth.

Many researchers draw on secondary analyses of test items and examine how specific text and item characteristics are linked to individuals’ performance on those test items [12, 18]. For this, it is necessary to analyze said texts and items as a first step. Text and item characteristics can be manually tagged, coded and counted. For instance, Shaftel et al. [19] created a coding list for raters in order to rate test items regarding 17 characteristics. However, this procedure might be error prone and time consuming. Thus, there are many natural language processing (NLP; see [20] for a detailed description) tools that automatize this process in order to save resources [21]. There are many tools researchers and practitioners can choose from depending on the linguistic analyses they are aiming for. For instance, there are many (online) tools for parts of speech tagging, such as the online demo version of the Stanford CoreNLP [22]. In case users are interested in readability indices, they can opt for ARTE [3], when calculating English texts and items or RATTE [4], when analyzing German texts or items. Users are even confronted with more tools to choose from, when they want to analyze rather objective text and item characteristics, such as sentence length or word count (e.g., Coh-Metrix [23]; the SiNLP [Simple Natural Language Processing Tool] [21]).

However, it might be time consuming having to use several tools at once, which all might have their own strengths but also shortcomings. For instance, most tools are only available in one language [1, 3, 23, 24]. More importantly, to the best of our knowledge, several developers provide no or only limited accuracy measures for the individual features implemented in their tools [4, 21], which makes it difficult to decide whether the results are reliable or not. With LATIC, we aim to provide users with a tool that resolves some of the shortcomings other tools have. In the following, we introduce our tool and investigate how accurate LATIC’s results are compared to human ratings. In the discussion, we will further compare LATIC’s features with other tools, highlighting advantages as well as limitations of the tool.

The software LATIC

Technical details

LATIC is a java application that is free and open source. LATIC can be used on Windows and Linux (e.g., Debian, Ubuntu) operating systems as well as on macOS to analyze texts and
items. The duration of text and item analyses depends on the performance of the computer and the extent of the analyses. In our test runs, the analyses usually took a few seconds only.

For some features (i.e. tagging parts of speech), LATIC relies on a NLP tool. For LATIC, we used a NLP tool named Stanford CoreNLP [5]. The Stanford CoreNLP [25] is Java application providing users with several linguistic annotations for any kind of text, including parts of speech tagging. We always incorporate the latest version of the Stanford CoreNLP [5]. When submitting this manuscript, version 4.4.0 was the newest.

**How to use LATIC.** LATIC is an intuitive and easy to use software that can be downloaded on https://download.latic.software. Users are able to use the software after installing LATIC (a step-by-step instruction is provided with the download). However, if users do not want to install LATIC, they can start it via the console. With using LATIC, data security is guaranteed since all analyses are performed on users' local computers.

Fig 1 depicts the user interface of LATIC (version 1.2.2). In order to analyze a text or an item, users first need to choose the language of the text or item on the upper right. Second, users must enter the text or item in the text field or upload a document containing the text or item. Third, the users choose the characteristics that shall be analyzed on the right. Finally, the users need to click on the “Analyze” button to display the results. This step might take a few seconds. If users want to save their results, they can do so by clicking on the “Save file” button and choosing an appropriate file format, such as .csv files. The “Delete” button enables users to delete the results. A more detailed example is uploaded on GitHub [26]: It demonstrates the analysis of a famous tale and includes step-by-step instructions (including screenshots) to obtain the results.

If users need help, they can click on the “Help” button on the top row. They are provided with the option to write an e-mail or to retrieve the documentation of the software [27], including step-by-step instructions and detailed information about every feature of LATIC (see next section).
In order to ensure correct results, users need to keep a few factors in mind. First, similar to other software, it is essential to use correct spelling for reliable results, especially with regard to the parts of speech tagging. Second, abbreviations and measuring units should be avoided since they tend to bias both the tagging and the calculation of item characteristics (e.g., readability indices, average word length). Third, cardinal numbers should be written as Arabic numbers if tagging parts of speech is relevant for the analysis of the item. To calculate other characteristics, such as the syllable count or a readability index, the cardinal numbers should be spelled out. Finally, it is important to note that the Stanford CoreNLP [5] tags every juxtaposition of characters separated by spaces or punctuation marks individually. Hence, for instance, Red Sea would be counted as two words, each getting the tag of a proper noun.

Features

LATIC allows users to analyze texts and items regarding their characteristics at different levels, i.e., the word, sentence and text level. Table 1 includes all features of LATIC. Due to the amount of features that LATIC provides, going into detail for every single feature, would go beyond the scope of this paper. In the documentation of LATIC [27], we present every single feature of LATIC in much more detail, including short explanations and examples.

First, LATIC enables the users to tag and count different parts of speech contained in a text or an item. The tagging of parts of speech is mostly based on the Stanford CoreNLP [5]. The only exceptions are the taggings of primary interjections and symbols in the French, German, and Spanish versions due to frequent errors in our test runs; these taggings are made by LATIC. The annotation in English is based on the Penn Treebank Tagset [28]; for French, German and Spanish the UD (version 2) tagset is used [29].

Furthermore, LATIC counts up the tags, thus providing the user with an absolute frequency of the parts of speech chosen for analysis. LATIC is also able to display the tags for every word or rather character juxtaposition. This feature enables the user to examine the tags given out by the Stanford CoreNLP [5].

Second, LATIC enables the user to analyze texts and items regarding other objective characteristics at all three levels, such as the average word and sentence length (see Table 1).

Third, LATIC’s features include the calculation of several readability indices (see Table 1). As noted above, traditional readability indices are heavily criticized [17] and we mostly agree with the critique. However, since the calculation of readability indices can be helpful for some research questions [15], we want to enable users to calculate them more easily. One must note that not every readability index is available in all languages because readability indices are not necessarily transferrable to other languages [30]. For instance, the Flesch Reading Ease [6] and the SMOG [16] for English texts and items needed to be adapted for the German language, resulting in the Flesch index [31] and the gSMOG [32]. Hence, we only implemented readability indices suitable for the respective languages.

The present study

In this study, we aim to first examine the performance of LATIC in terms of accuracy as well as time saving compared to analyses performed by human raters. For this, we aimed to focus on three aspects. During the development of LATIC, the implemented features were repeatedly tested and bugs were fixed immediately when noticed. However, there were two exceptions: The part of speech tagging by means of the Stanford CoreNLP 4.4.0 [5] as well as the syllable count. No errors occurred during our multiple test runs regarding all other features (such the counting of parts of speech and the calculation of text and item characteristics). Thus, we particularly wanted to investigate the accuracy of (1) the parts of speech tagging and (2) the
Furthermore, we aimed to investigate how the performance of LATIC compares to humans’ performance in terms of accuracy as well as time. Examining these aspects, we focus on the English and German language.

| Level          | Text and item characteristics | Available in |
|----------------|--------------------------------|--------------|
| Word           | Adjectives                     | ✓            |
|                | Adpositions                    | ✓            |
|                | Adverbs                        | ✓            |
|                | Cardinal numbers               | ✓            |
|                | Conjunctions                   | ✓            |
|                | Determiners                    | ✓            |
|                | Existential there              | ✓            |
|                | Infinitive to                  | ✓            |
|                | Interjections (only primary)   | ✓            |
|                | List item markers              | ✓            |
|                | Modals                         | ✓            |
|                | Nouns                          | ✓            |
|                | Particles                      | ✓            |
|                | Possessive endings             | ✓            |
|                | Prepositions                   | ✓            |
|                | Pronouns                       | ✓            |
|                | Proper nouns                   | ✓            |
|                | Punctuation marks              | ✓            |
|                | Symbols                        | ✓            |
|                | Tenses                         | ✓            |
|                | Unknown or uncertain           | ✓            |
|                | Verbs                          | ✓            |
|                | Word length (characters)       | ✓            |
|                | Word length (syllables)        | ✓            |
| Sentence       | Number of sentences            | ✓            |
|                | Sentence length (characters w/o spaces) | ✓ |
|                | Sentence length (characters w/ spaces) | ✓ |
|                | Sentence length (syllables)    | ✓            |
|                | Sentence length (words)        | ✓            |
| Text           | Automated readability index    | ✓            |
|                | Coleman-Liau index             | ✓            |
|                | Flesch-Kincaid grade level     | ✓            |
|                | Flesch reading ease            | ✓            |
|                | gSMOG                          | ✓            |
|                | Gunning Fog index              | ✓            |
|                | Lexical diversity              | ✓            |
|                | LIx                            | ✓            |
|                | SMOG                           | ✓            |
|                | Syllable count                 | ✓            |
|                | Tagging the whole item         | ✓            |
|                | Wiener Sachtextformel          | ✓            |
|                | Word count                     | ✓            |

https://doi.org/10.1371/journal.pone.0277250.t001
Method

Parts of speech tagging

In order to test the accuracy of the tagging provided by the Stanford CoreNLP 4.4.0 [5], we compared the tokens given out by LATIC with the tokens of two well-known corpora (see [27] for more details). For the English language, we used the MULTTEXT-East 4.0 corpus [33]. The corpus was manually tagged and consists of the book “1984” by George Orwell [34]. For the German language, we used the TIGER corpus 2.2 [35]. The corpus was tagged semi-automatically and consists of newspaper articles [36].

For evaluating the accuracy, we used around 10,000 tokens of each corpus. These tokens were analyzed with LATIC. However, to compare the accuracy of LATIC in a reasonable way, we needed to take a few steps. First, we deleted all words that were misspelled since a right spelling of words is essential for tagging the parts of speech correctly. Second, we matched the respective tagsets to the ones used by the Stanford CoreNLP 4.4.0 [5]. Pronouns, for instance, are tagged in much more detail in the MULTTEXT-East 4.0 corpus [37] than in the tagsets used by the Stanford CoreNLP [5]. Third, we examined how many tokens were tagged correctly by LATIC by means of the Stanford CoreNLP 4.4.0 [5].

Calculation of syllable count

In the English language, we used a script [38] to count the syllables. First, we tested the script with $N = 9,107$ common words (https://en-syllables.latic.software/). In a second step, we optimized the results by adapting the script: We created a list of words that were counted incorrectly and implemented the correct syllable count for each of those words [39].

In the German language, we created an algorithm to count the syllables. First, we tested this algorithm with $N = 9,524$ common words according to the Leipzig Corpora Collection [40]. Again, in a second step, we optimized the algorithm by including (1) additional rules to correctly count the syllables, (2) a list of words, in which the syllables were not correctly counted, and (3) Anglicisms and Gallicisms that are common in German.

Comparison with human performance

We chose five openly available science items that were used in the PISA study (Programme for International Student Assessment) in 2015 [41]: (1) the introductory text of item S656, (2) S656Q01, (3) S656Q02 as well as (4) the introductory text of item S641, and (5) the item S641Q01. The chosen items ranged from 22 to 101 words and did not include any illustrations.

We asked four people to analyze the items regarding the selected text and item characteristics. The four raters were two student assistants with little experience in linguistic coding, one foreign language assistant with decent experience in linguistic coding as well as the first author of this paper who has a lot of experience in linguistic coding. The first three raters were given an instruction on how to analyze text and item characteristics by the first author. Especially the two student assistants received further explanations and support due to their little experience in coding.

We randomly selected thirteen text and item characteristics that the raters should analyze (see Table 2). The raters were asked to (1) give their ratings regarding the characteristics, (2) state how confident they were in their results (ranging from 1 [not at all] to 5 [very confident]), and (3) to measure how long it took for them to analyze each item.

Ethical approval

An ethics committee approval was not needed due to solely working with entered text or items in LATIC. Sensitive data, such as from vulnerable people, were not used in developing the
All four raters of legal age gave their consent in participating in the study.

Results

Parts of speech tagging
Regarding the plain comparison of tokens, LATIC correctly tagged $n = 8,651$ out of all $N = 9,989$ (86.61%) in the English language, and $n = 9,093$ out of all $N = 9,997$ (90.96%) in the German language. As noted above, some modifications needed to be made to ensure a reasonable comparison of the tokens. After these modifications, $n = 9,048$ out of the remaining $N = 9,755$ (92.75%) of the tokens were correctly tagged by the Stanford CoreNLP 4.4.0 [5] in the English language; in the German language, $n = 9,271$ out of the remaining $N = 9,879$ (93.85%) of the tokens were correct. Finally, we analyzed the most frequent errors in tagging by examining the first 5,000 tokens. In both languages, the differentiation between determiners and pronouns seemed to be the most challenging (English: 37.84% of all errors; German: 21.74% of all errors). However, this was to be expected since the differentiation between these two word classes is not always clear [29]. In the English language, the distinction of adjectives and the tagging of (1) prepositions and subordinating conjunctions (11.28% of all errors), and (2) all types of verbs (8.52% of all errors) were by far the second and third most frequent errors. In the German language, the second and third most frequent errors were the distinction between nouns and proper nouns (13.71% of all errors) and the distinction between adjectives and adverbs (9.36% of all errors).

Calculation of syllable count
In the English language, the syllables of $n = 8,723$ words (95.78%) were counted correctly. After adapting the script by Wormer (2021), the syllables of all $N = 9,107$ test words (100%) were correctly counted. In the German language, the syllable count of $n = 9,190$ words (96.49%) was correctly calculated. After optimizing our algorithm, the syllable count was correct for $n = 9,522$ test words (99.98%).

Comparison with human performance
All results by the four raters as well as LATIC’s results are openly available on https://rating-results.latic.software/. The means of the raters’ ratings as well as LATIC’s results are depicted in Table 2. In terms of accuracy, the raters felt confident that their ratings were accurate ($M = 4.27, SD = 0.55$). In fact, conducting a Mann-Whitney-U-test, there are no significant differences between the coders’ and LATIC’s results, $U = 2081.50, p = .885$. We also tested whether the overall ratings of each individual rater differed from LATIC’s results by conducting further Mann-Whitney-U-tests. All differences were non-significant ($p > .570$).

Despite the differences not being significant, LATIC still achieves higher accuracy than the raters regarding certain text and item characteristics, such as calculating the lexical diversity [42] and the readability index LIX [7]. This can be mostly traced back to human raters’ calculation errors, which LATIC as a technical software does not make.

The interrater-reliability (Krippendorff’s $\alpha$) [43] varies tremendously between the different text and item characteristics. It can be positively mentioned that the raters reach (very) good reliabilities (Krippendorff’s $\alpha \geq 0.80$) for more than half of the text and item characteristics (e.g., number of sentences, word count). However, it becomes apparent that not all ratings made by humans are suitable to work with since some interrater reliabilities are very low (e.g., pronouns, LIX).
In total, the four raters needed $M = 4,374.80$ seconds (~72 minutes; $SD = 1,846.25$) to analyze all five items regarding the thirteen text and item characteristics. For the analyses of each item, the four raters used $M = 2.60$ ($SD = 0.55$) different tools, which they used $M = 7.80$ ($SD = 0.45$) times. The first author further analyzed the five items by means of LATIC and finished the analyses within 80 seconds, including starting the software and choosing the text and item characteristics on the right sidebar (see section “How to Use LATIC”).

**Discussion**

As noted above, there are a few tools and web applications used by researchers to automatize the process of analyzing text and item characteristics. However, these tools and applications might only be partially suitable for envisioned purposes. Thus, in this article, we introduced the new tool LATIC which combines many linguistic features, such as parts of speech tagging, the calculation of objective text and item characteristics as well as readability indices. On the basis of the Stanford CoreNLP [5], users can analyze texts and items in English, French, German, and Spanish. Drawing on our results, LATIC achieves very high performance in terms of accuracy as well as time saving compared to human raters. Thus, LATIC provides reliable results, while also saving users’ resources.

**Strengths of LATIC**

In research, it is important to work with reliable results in order to draw plausible and accurate conclusions. Thus, it is essential to work with tools and software that provide researchers with...
such reliable results. Unfortunately, to the best of our knowledge, many authors also fail to provide measurements of accuracy for their tools or rather for specific features of their tools (for example [4]). Our article shows that users can analyze linguistic text and item characteristics with high accuracy in a short time with LATIC. The tagging provided in LATIC by means of the Stanford CoreNLP 4.4.0 [5] reaches very good accuracy in our analyses, which corroborates prior research. In other evaluation studies, the Stanford CoreNLP [5] usually reaches one of the best accuracy rates compared to other parts of speech taggers [44, 45]. Furthermore, the accuracy of the syllable count is also high, especially when adapting the script and algorithm to optimize the syllable count feature. Finally, comparing LATIC’s performance with human raters, the results of our analyses showed that LATIC reaches similar accuracy as human raters. At this point, it is important to note that the interrater-reliability varies tremendously for some linguistic text and item characteristics suggesting that a tool like LATIC might provide more adequate rating results than human raters. In sum, LATIC provides reliable results while saving the users valuable resources, such as time and effort, since the analyses usually only takes (milli-) seconds.

In terms of scientific aspects, many authors traditionally tend to manually tag, code and count text and item characteristics, such as word count [9] or prepositions [19]. While doing this, the human raters might fall back on several different tools, which was the case in our study. This might be due to some web applications offering only a limited number of features (e.g., T.E.R.A [46]) or even just one (e.g., LIX calculator [47]). LATIC provides the users with the advantage of combining many features in one software that are essential for evaluating the readability of texts and items. As noted above, LATIC as a tool offers users to automatically tag and count parts of speech. Furthermore, LATIC calculates further characteristics as well as traditional readability indices, which are regularly used in research [10, 19, 48].

In terms of breadth of features, most tools and applications only analyze texts or items in one specific language (e.g., ARTE [3]). LATIC offers the linguistic analyses of texts and items in four languages (English, French, German, and Spanish). Further languages can be implemented into LATIC, if suitable natural language processing tools are available. For instance, besides the languages already implemented into LATIC, the Stanford CoreNLP further supports tagging parts of speech in the Arabic, Hungarian, Italian, and Chinese languages [25]. However, since none of the authors is fluent in neither of those languages, these could not be implemented (yet).

Regarding usability aspects, some web applications are not always available online (Coh-Metrix [46]; Stanford CoreNLP demo [22]), require a registration (e.g., CTAP [49]; Text Inspector [24]), or are not free to use (e.g., LIWC [50]; Text Inspector [24]). LATIC is open source and free to use. Thus, the code is openly available [39] and can be audited and modified. Furthermore, users can request features to be implemented into LATIC and may report bugs under https://github.com/florianklueckmann/LATIC/issues or by sending an e-mail to hello@latic.software. It is also of advantage that the interface is user-friendly and intuitively usable, thus diminishing long training periods. All features are documented in detail [27], which provides users with essential information on how the results of LATIC are formed. LATIC also enables users to work with the results by allowing users to save and import the results into statistical analysis software, if needed.

Limitations

Despite the strengths of LATIC, a few limitations need to be discussed. First, in order to obtain the most reliable results possible, users need to follow a few instructions. Similar to other tools and software, this includes ensuring the correct spelling and avoiding abbreviations (see
above). This means that users have a high responsibility to ensure the entered texts and items are examined and—if necessary—modified accordingly. This means that, for instance, texts written by students can only be analyzed regarding their characteristics, if the spelling is corrected beforehand. However, it might be the case that researchers or practitioners are especially interested in these errors. In this case, we would recommend using the software GAMET (Grammar and Mechanics Error Tool) [51]. The software enables users to locate structural as well as mechanical errors within an English text, which is the focal point of the tool. Parts of speech tagging or the calculation of text and item characteristics, however, cannot be executed.

Second, some linguistic characteristics prominent in readability research, such as cohesion [52] or phrases [18], cannot be analyzed in LATIC. However, there are a number of tools enabling users to analyze their English texts and items regarding cohesion. For instance, the tool TAACO [1] estimates over 150 indices of global and local cohesion. While TAACO provides many indices of cohesion, other text and item characteristics, such as readability indices or the syllable count, cannot be calculated. In case, users want to analyze phrases, we recommend the tool TAASSC (Tool for the Automatic Analysis of Syntactic Sophistication and Complexity) [53, 54]. While TAASC focuses on estimating the syntactic sophistication as well as complexity of English texts, parts of speech tagging as well as the calculation of the syllable count and readability indices are not available.

Finally, when developing LATIC, we decided to implement the Stanford CoreNLP [5] rather than other NLP tools due to its high accuracy rates in prior research [44, 45]. However, one must note that the Stanford CoreNLP [5] seems to perform relatively well on generic text and item types, but not in significantly different text types. For instance, other parts of speech taggers outperform the Stanford CoreNLP [5] in social media texts [55] or ReadMe documents of software [56]. The main intent was to provide educators and (readability) researchers the opportunity to analyze linguistic text and item characteristics with LATIC. Thus, we opted for an NLP tool that performed well on rather formal text and item types found, for instance, in the educational context.

Conclusion
The analyses of linguistic text and item characteristics might not only be necessary for investigating research questions regarding language in texts and items, but also for practitioners, such as teachers and educators. Traditionally, texts and items are analyzed manually [9, 19], which is very time consuming. LATIC is a user friendly and free to use java application enabling users to analyze texts and items in four languages (English, French, German, and Spanish) regarding their linguistic characteristics. LATIC offers features at the word, sentence and text level including the tagging and counting of parts of speech and the calculating of traditional readability indices. All features were tested thoroughly and reach remarkable accuracy.

Acknowledgments
We would like to thank Judith Keinath for the editorial assistance with this article.

Author Contributions
Conceptualization: Nadine Cruz Neri, Florian Klückmann, Jan Retelsdorf.
Data curation: Nadine Cruz Neri.
Formal analysis: Nadine Cruz Neri.
Methodology: Nadine Cruz Neri, Florian Klückmann.

Software: Florian Klückmann.

Validation: Nadine Cruz Neri.

Visualization: Nadine Cruz Neri.

Writing – original draft: Nadine Cruz Neri.

Writing – review & editing: Jan Retelsdorf.

References

1. Crossley SA, Kyle K, McNamara DS. The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. Behavior Research Methods 2016; 48(4): 1227–1237. https://doi.org/10.3758/s13428-015-0651-7 PMID: 26416138

2. Dascalu M, Gutu-Robu G, Ruseti S, Paraschiv IP, Dessus P, McNamara D, et al. ReaderBench: A multi-lingual framework for analyzing text complexity. In: Lavoué É, Drachsler H, Verbert K, Broisin J, Pérez-Sanagustin M, editors. Data Driven Approaches in Digital Education. Cham: Springer; 2017. p. 495–499.

3. Choi JS, Crossley SA. Readability assessment tool for English texts [Internet]. EasyChair [Preprint no. 6190]. 2021 [cited 2022 14 Sep]; 12 p. Available from: https://easychair.org/publications/preprint/pgb8.

4. Wild J, Pissarek M. [RATTE. Regensburg analysis tool for texts] RATTE. Regensburger Analystool für Texte. Version 2.0 [software]. 2022 Aug 25 [cited 2022 14 Sep]. Available from: https://www. http://ratte.lese didaktik.net/

5. Manning CD, Surdeani M, Bauer J, Finkel J, Bethard SJ, McClosky D. The Stanford CoreNLP Natural Language Processing Toolkit. In: Bontcheva K, Zhu J, editors. Proceedings of 52nd Meeting of the Association for Computational Linguistics: System Demonstrations; 2014, Jun 23–24; Baltimore, United States. United States: The Association for Computational Linguistics. p. 55–60.

6. Flesch R. A new readability yardstick. Journal of Applied Psychology 1948; 32(3): 221–233. https://doi.org/10.1037/h0057532 PMID: 18867058

7. Björnsson C-H (1968). [Readability] Läsbarhet. Kopenhagen: Gad.

8. Lesaux NK, Lipka O, Siegel LS. Investigating cognitive and linguistic abilities that influence the reading comprehension skills of children from diverse linguistic backgrounds. Reading and Writing: An Interdisciplinary Journal 2006; 19(1): 99–131. https://doi.org/10.1007/s11145-005-4713-6

9. Cruz Neri N, Guill K, Retelsdorf J. Language in science performance: Do good readers perform better? European Journal of Psychology of Education 2021; 36(1): 45–61. https://doi.org/10.1007/s10212-019-00453-5

10. Cruz Neri N, Retelsdorf J. Do students with specific learning disorders with impairments in reading benefit from linguistic simplification of test items in science? Exceptional Children. https://doi.org/10.1177/00144029221094049

11. Abedi J, Lord C. The language factor in mathematics tests. Applied Measurement in Education 2001; 14(3): 219–234. https://doi.org/10.1207/S15324818AME1403_2

12. Cruz Neri N, Wagner J, Retelsdorf J. What makes mathematics difficult for adults? The role of reading components in solving math items. Educational Psychology 2021; 41(9): 1199–1219. https://doi.org/10.1080/01443410.2021.1964438

13. Stevens SS, Stone G. Psychological writing, easy and hard. American Psychologist 1947; 2(7): 230–235. https://doi.org/10.1037/h0062792 PMID: 20254703

14. White S. Mining the text: 34 text features that can ease or obstruct text comprehension and use. Literacy Research and Instruction 2012; 51(2): 143–164. https://doi.org/10.1080/19388071.2011.553023

15. King D, Burge B. Readability analysis of PISA 2012 mathematics, science and reading assessments. Slough: National Foundation for Educational Research; 2015.

16. McLaughlin GH. SMOG grading—a new readability formula. Journal of Reading 1969; 12(8): 639–646.

17. Lenzner T. Are readability formulas valid tools for assessing survey question difficulty? Sociological Methods & Research 2014; 43(4): 677–698. https://doi.org/10.1177/0049124113513436

18. Haag N, Heppt B, Stanat P, Kuhl P, Pant HA. Second language learners' performance in mathematics: Disentangling the effects of academic language features. Learning and Instruction 2013; 28: 24–34. https://doi.org/10.1016/j.learninstruc.2013.04.001
19. Shaftel J, Belton-Kocher E, Gliasnapp D, Poggio J. The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. Educational Assessment 2006; 11(2): 105–126. https://doi.org/10.1207/s15326977ea1102_2
20. Chowdhury GG. Natural language processing. Annual Review of Information Science and Technology 2005; 37(1): 51–89. https://doi.org/10.1002/aris.1440370103
21. Crossley SA, Allen K, Kyle K, McNamara DS. Analyzing discourse processing using a simple natural processing tool. Discourse Processes 2014; 51(5–6): 511–534. https://doi.org/10.1080/0163853X.2014.910723
22. Stanford NLP Group. Core NLP version 4.4.0 [Internet]. [place unknown]: [publisher unknown]; 2022 [cited 2022 Sep 14]. Available from: https://stanfordnlp.github.io/CoreNLP/human-languages.html.
23. Stanford NLP Group. Using CoreNLP on other human languages [Internet]. [place unknown]: Stanford NLP Group; 2020 [cited 2022 Sep 14]. Available from: https://github.com/florianklueckmann/LATIC.
24. Bamberger R, Vanecek E. [Reading-comprehension-learning-writing. The difficulty levels of texts in the German language] Lesen-Verstehen-Lernen-Schreiben. Die Schwierigkeitsstufen von Texten in deutscher Sprache. Wien: Jugend und Volk; 1984.
25. Erjavec T, Barbu A-M, Derzhanski I, Dimitrova L, Garabik R, Ide N, et al. MULTTEXT-East “1084” annotated corpus 4.0. Slovenian resource repository CLARIN [Data set]. Jožef Stefan Institute: Slovenia; 2010, [2022 Sep 19]. Available from: https://www.clarin.si/repository/xmlui/handle/11356/1043.
26. Erjavec T. MULTTEXT-East: Morphosyntactic resources for central and eastern European languages. Language Resources & Evaluation 2012; 46: 131–142. https://doi.org/10.1007/s10579-011-9174-8
27. Brants T, Plaehn O. Interactive Corpus Annotation. In: Gavrilidis M, Carayannis G, Markantonatou S, Piperidis S, Steinhauser G, editors. Proceedings of the second international conference on language resources and evaluation (LREC-2000); 2000 May 31–Jun 2; Athens, Greece. Luxemburg: European Language Resources Association; 2000.
28. Ide N, Priest-Dorman G, Erjavec T, Varadi T. MULTTEXT-East morphosyntactic specification, version 4. English specifications [Internet]. [place unknown]: Natural Language Server; 2009 Oct 6 [cited 2022 Sep 14]. Available from: http://nl.ija.si/ME/Vault/V4/msd/html/msd-en.html.
29. Wormer T. Syllable; 2021 [cited 2022 Sep 19]. Database: GitHub repository [Internet]. Available from: https://github.com/words/syllable
30. Knückmann F. LATIC; 2021 [cited 2022 Sep 19]. Database: GitHub repository [Internet]. Available from: https://github.com/florianklueckmann/LATIC.
31. Latic – A linguistic analyzer
41. OECD. Try PISA 2015 test questions [Internet]. [place unknown]: OECD; [cited 2022 Sep 19]. Available from: https://www.oecd.org/pisa/test/PISA%202015%20MS%20-%20Released%20Item%20Descriptions%20Final_English.pdf.

42. Johansson V. Lexical diversity and lexical density in speech and writing: A developmental perspective. Working Papers 2009; 53(2008): 61–79.

43. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. Communication Methods and Measures 2007; 1(1): 77–89. https://doi.org/10.1080/19312450709336664

44. Giesbrecht E, Evert S. Is part-of-speech tagging a solved task? An evaluation of POS taggers for the German web as corpus. In: Alegria I, Leturia I, Sharoff S, editors. Proceedings of the 5th Web as Corpus Workshop (WAC5); 2009 Sep 7; San Sebastian, Spain. San Sebastian: Elhuyar Fundazioa. p. 27–36.

45. Tian Y, Lo D. A comparative study on the effectiveness of part-of-speech tagging techniques on bug reports (pp. 570–574). In: Guéhéneuc Y-G, Adams B, Serebrenik A, editors. 2015 IEEE 22nd International Conference on Software Analysis, Evolution, and Reengineering (SANER); 2015 Mar 2–6; Montreal, Canada. New Jersey, United States: Institute of Electrical and Electronics Engineers. p. 570–574.

46. McNamara DS, Graesser A, Cai Z, Dai J. Coh-Metrix common core T.E.R.A. Version 1.0 [software]. 2012 [cited 2022 13 Sep]. Available from: http://cohmetrix.com/.

47. Lenhard W, Lenhard A. [Calculation of the LIX readability index by Björnson] Berechnung des Lesbarkeitsindex LIX nach Björnson [Internet]. Dettelbach: Psychometrica; 2014–2022 [cited 2022 13 Sep]. Available from: http://www.psychometrca.de/lix.html.

48. Brysbaert M, Sui L, Duyck W, Drix N. Improving reading rate prediction with word length information: Evidence from Dutch. Experimental Psychology Society 2021; 74(11): 1–6. https://doi.org/10.1177/17470218211017100 PMID: 33910411

49. Chen X, Meurers D. CTAP: A web-based tool supporting automatic complexity analysis. In: Brunato D, Dell’Orletta F, Venturi G, François T, Blache P, editors. Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity; 2016 Dec 11–17; Osaka, Japan. Japan: The COLING 2016 Organizing Committee; 2016. p. 113–119.

50. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. Austin: University of Texas at Austin; 2015.

51. Crossley SA, Bradfield F, Bustamante A. Using human judgments to examine the validity of automated grammar, syntax, and mechanical errors in writing. Journal of Writing Research 2019; 11(2): 251–270. https://doi.org/10.17239/jowr-2019.11.02.01

52. Hall SS, Kowalski R, Paterson KB, Basran J, Filik J, Maltby J. Local text cohesion, reading ability and individual science aspirations: Key factors influencing comprehension in science classes. British Educational Research Journal 2014; 41(1): 122–142. https://doi.org/10.1002/berj.3134

53. Kyle K. Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication. Doctoral dissertation, Georgia State University. 2016. Available from: http://scholarworks.gsu.edu/alesl_diss/35.

54. Lu X. Automatic analysis of syntactic complexity in second language writing. International Journal of Corpus Linguistics 2010; 15(4): 474–496. https://doi.org/10.1075/ijcl.15.4.02lu

55. Pinto A, Oliveira HG, Oliveira Alves A. Comparing the performance of different NLP toolkits in formal and social media text. In: Mernik M, Leal JP, Oliveira HG, editors. 5th Symposium on Languages, Applications and Technologies; 2016 Jun 21; Dagstuhl, Germany. Dagstuhl: Leibniz-Zentrum für Informatik. p. 3:1–3:16.

56. Al Omran FNA. Treude C. Choosing an NLP library for analyzing software documentation: A systematic literature review and a series of experiments. In: [editors unknown]. 2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR); 2017 May 20–21; Buenos Aires, Argentina. New Jersey, United States: Institute of Electrical and Electronics Engineers. p. 187–197.