

SECTION: PHILOLOGICAL SCIENCES.

SEKCJA: NAUKI FILOLOGICZNE.

How to cite: Kapeta, C. (2024). Basic elements for determining the language level and the Readability grade in text production. *International Conference on Science, Innovations and Global Solutions*. (pp. 222-229). Futurity Research Publishing. <https://futuraity-publishing.com/international-conference-on-science-innovations-and-global-solutions-archive/>

Basic elements for determining the language level and the Readability grade in text production

Chrysovalantou Kapeta

PhD, Researcher, Department of Italian Language, Aristotle University of Thessaloniki, Greece

Accepted: July 11, 2024 | **Published:** July 24, 2024 | **Language:** English

Abstract: In order to measure Readability, 316 texts from the state certificate for Italian language in Greece were selected (May 2015 & November 2016). The texts were manually digitized in Word 2010 format and measured through the Read-IT tool. The data processing was carried out using SPSS software 24. Greek candidates at level B produce texts in Italian depending on features interwoven with vocabulary, grammar and syntax. The aim was to form the basis for further research in other language levels (A & C) and different languages which would probably provide important outcomes for language learning by foreign speakers.

Keywords: criteria, evaluation, measurement, readability, grammar, syntax, vocabulary, level, validity

Introduction

Appropriate tools lead to valid and fair measurement, processing and evaluation of texts produced by foreigners in a different language (Elder & Harding, 2008). In addition, this data is very important in order to be able to create exam instructions and tests depending on language skills (in this case B1 and B2 based on the Common European Framework of Reference for Foreign Languages) and level of difficulty, i.e. the readability grade (Lenzner, 2014). The discovery of such data is necessary for language levels (A1-C2) and for different languages in order to create more advanced text evaluation software (Beacco, 2017). Furthermore, the two largest universities EKPA and the University of Aristotle in Greece already use a very important measurement tool called *trat.exe* (Klonis, 2019), with which measurements of the level of difficulty for the preparation of Italian language exams of the State Certificate of Language Proficiency in Greece (KPG). By measurement we mean the use of readability formulas that use algorithms to calculate the number of word and sentence lengths, verbs, nouns, adjectives, etc. and thus measure the level of difficulty and foreign language level of a written text.

This newly developed software, or *trat.exe* (Klonis, 2019), is a reliable and valid digital tool for assessing the difficulty of Italian language tests. This software is used to develop test topics of various levels of language proficiency certification by measuring texts depending on the foreign language level and level of difficulty. In other words, it measures and calculates the difficulty level of the tests entered in Italian. In particular, the web software created for this purpose has the following functions:

- The reader can enter their text and the difficulty level results will be extracted by the software.
- User can see some additional features of the system such as: B. the relevance between number of words, syllables, characters (question or exclamation mark, period, etc.), sentences, syllables per word, words per sentence.
- It is also possible to read texts on an entire website by entering the corresponding electronic address.
- Finally, it is possible to measure the readability of doc, docx and PDF files when these files are uploaded to the software (Klonis, 2019).

Very interesting variables have been arrived at in order to be able to develop valid (validity) and fair (fairness) tests (Elder & Harding, 2008).

Analysis of recent research and publications

By "Legibility" in English or "Leggibilità" in Italian we mean the form of writing a character or a word via the size (the font if we talk about text in digital form like in Word) (MacKay et al., 2010), the

contrast of the characters. These parameters are related to the level of difficulty in reading a word, sentence or an entire text (Zuffi et al., 2007). In other words, readable texts are easier to read in a shorter time than other unreadable texts, which often require more effort and time to be understood by the reader.

Readability in English (Crossley et al., 2017) means the set of elements in a text that affect reading. In short, the readability criteria, related to the understanding of the text, offer the possibility of being able to measure the text. On the other hand, the choice of these criteria depends on the needs and characteristics of the reader (Dale & Chall, 1949).

Formulation of the purpose of the article

This study aims at an accurate and objective text evaluation depending on the language level. It would also be very interesting to compare the same elements in exams such as the Greek certificate from different countries, so that new data can be obtained in the area of foreign language learning. A general international database would lead to results, as it would provide more objective tests for foreigners in learning foreign languages and could also find out new information about the level of difficulty per country in order to develop new software.

Presentation of the main material (research results)

Various results have emerged from this research and the following have been added and reviewed:

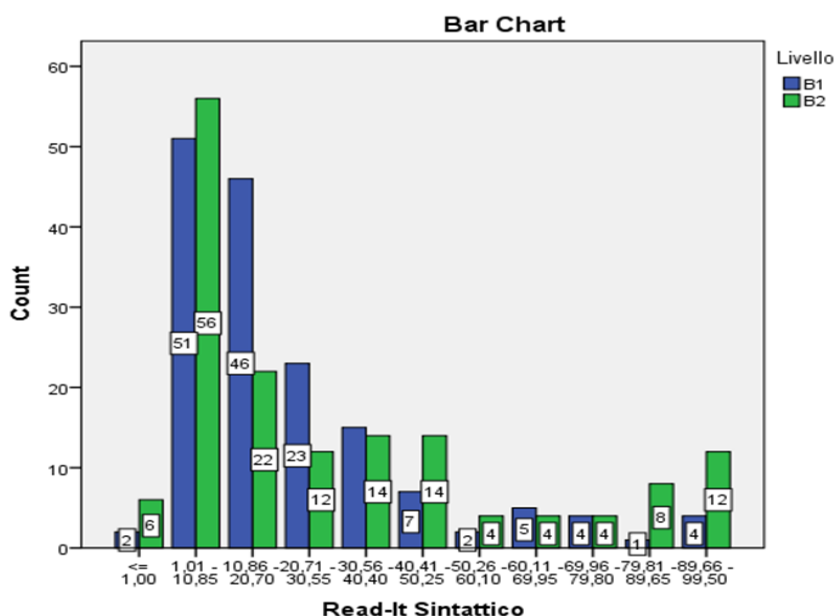
1. The criteria depending on the language level of written language production.
2. The criteria for the level of readability.
3. The advantages of readability for the Italian language, thereby creating a new innovative tool for texts, such as the Read-IT tool (Dell'Orletta, Montemagni & Venturi, 2011).
4. The importance of digital tools and their relationship to a fair assessment of Italian texts produced by Greeks and not by native Italian speakers.
5. In the following tables and graphs, important data such as the ability of Greek users of Italian in relation to the use of various features by foreign language level are observed.

Figure 1 shows us that the majority (107 texts, i.e. 56 at level B2 and 51 at level B1) only achieve 1.01-10.85%. It can be concluded that the Greek candidates seem to have difficulties in using syntactic and grammatical phenomena in the production of text in Italian. Although no text is 100% perfect, this does not mean that these texts lack correct use of syntactic and grammatical elements. This can be seen in a larger range of percentage points as one observes different texts at both levels

throughout the graph. In contrast, serious spelling errors were observed in many B1 and B2 texts because the texts were created by Greeks (e.g. confusion of words in other foreign languages such as English, German, Spanish or French). In addition, the grammar lacks the correct verb conjugation according to person or tense, and subordinate clauses are also missing in some texts. These facts are probably one of the reasons that help us distinguish a lower level of difficulty in the texts. On the other hand, the results show us that the authors of these texts are not Italian and are not bilingual either. The results for graded level B (B1 + B2) are therefore quite satisfactory and positive.

Figure 1

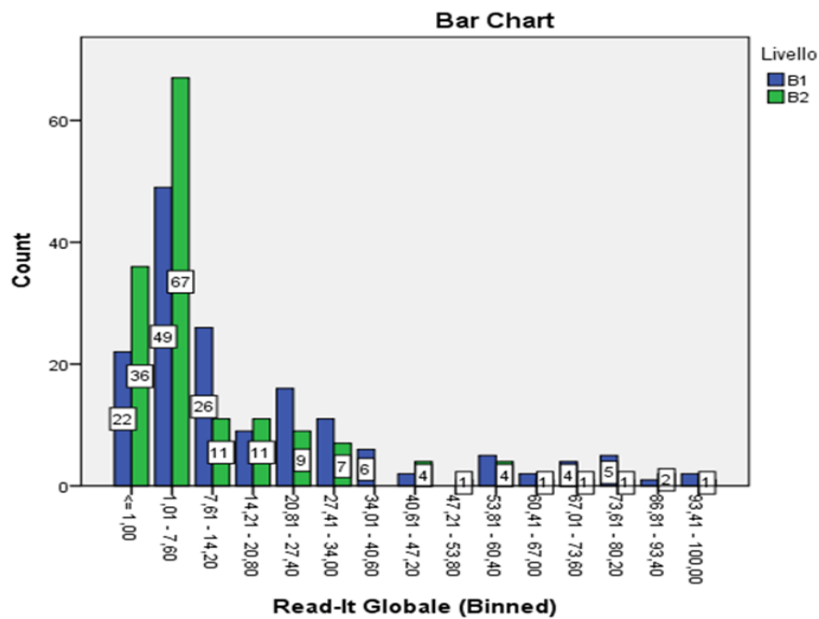
Display of the results of the READ-IT Sintattico index (level of difficulty for syntax) by foreign language level B1 or B2 (Livello)



In figure 2, the largest percentage of texts in terms of difficulty correspond to the foreign language level between 1.01% and 7.60%, with 67 texts at level B2 and 49 at level B1. It is important to note that the closer the percentage is to 100%, the greater the difficulty level appears to be. Therefore, in the present graphic we see a large volume of texts up to around the average, or up to 34%. This result shows that the texts created by Greek users are satisfactory.

Figure 2

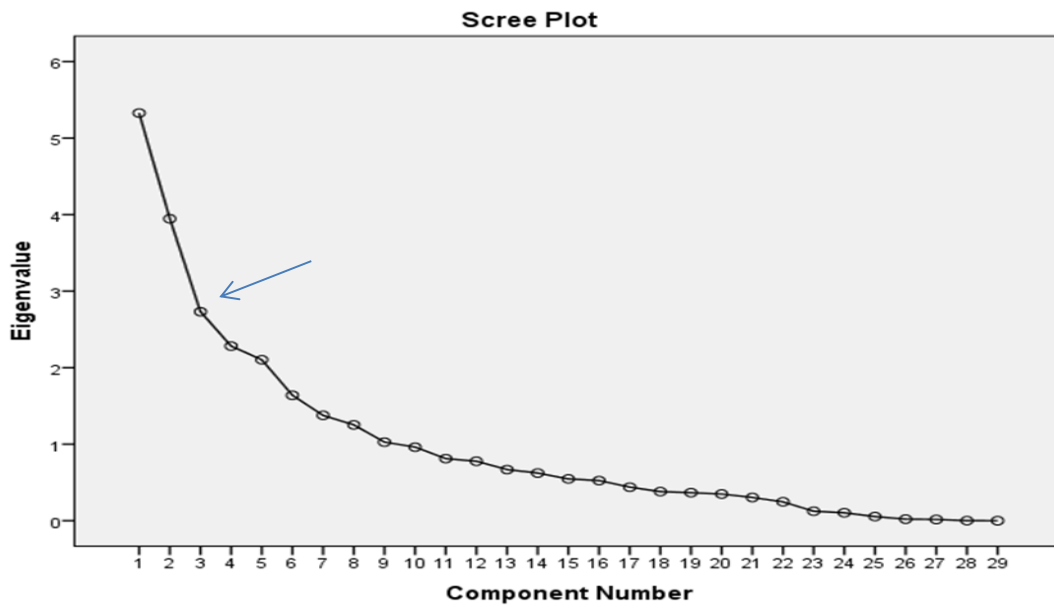
Display of the results of the level of difficulty (Read-IT Globale) per language level B1 or B2 (Livello)



After the factor analysis process, the so-called scree plot is created, through which new factors or components are discovered. For level B2, the first component (or the first new factor after the factor analysis) is the highest (5.32%). This is followed by the second (3.95%) and the third (2.73%). The possible variables that make up the first new factor that emerges from the level B2 data are word length (94%), sentence count (93%), and Gulpease index (90%). The Gulpease index refers to the index that measures the level of difficulty depending on whether a text is easy or difficult to read. The closer a text reaches 100%, the lighter it is. In this case the author uses, e.g. fewer verbs, nouns, subordinate clauses, spelling, conjunctions, adjectives, specific vocabulary where necessary (Figure 3):

Figure 3

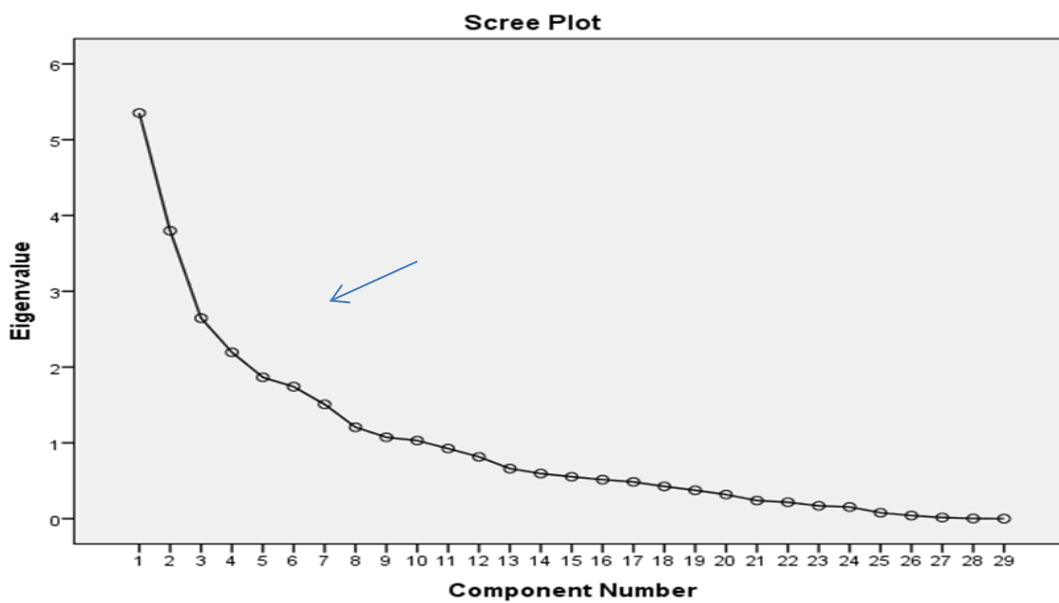
Scree plot of the first three new components of the variable for foreign language level B2



The first component is the highest (5.35%) for the B1 level texts (graph 4). This is followed by the second (3.80%) and the third (2.65%). The possible variables that make up the first new factor emerging from the data for level B1 are basic vocabulary (96%), word length (92%) and sentence number (93%).

Figure 4

Scree plot of the first three new components of the variable for foreign language level B1



Conclusions

From the new data, the following observations emerge regarding the main questions of this study: 1. The language level is related to the level of difficulty, if the first is low, the second is higher, and vice versa (there is an inversely proportional relationship between the two parts). 2. New factors emerging, or criteria that contribute to increasing the level of difficulty, located at both language levels (B1 and B2), and also include: basic vocabulary, word length, number of sentences and Gulpease index . 3. Apart from these elements, this study focuses on other factors that lead to either negative or positive results in a text production such as verbs, nouns, adjectives, coherence, and spelling. It is also noteworthy that factors such as nouns, adjectives or verbs seem to have less influence on the final result of the research, while on the contrary, conjunctions, for example, play a special role. Verbs (and sometimes adjectives) are typically more difficult for foreign language learners to handle than rigid conjunctions due to their inflectional properties. Many Greeks learn English as a second language and thus many words are confused from one language to another. This results in many language and spelling errors, as many words are written in English.

In conclusion, new factors differentiate each text according to language level and increase the level of difficulty. All these results mean that these important components may be included in future research in the field of linguistics and the development of innovative software such as the READ-IT tool (Dell'Orletta et al., 2014: 163-164) or trat.exe.

References

- Beacco, J.-C. (2017). *Il Quadro Comune Europeo di Riferimento per le Lingue - Un manuale Contributo all'insegnamento delle Lingue e al Piano Trentino Trilingue*. Trento: IPRASE-Istituto Provinciale per la Ricerca e la Sperimentazione Educativa, Working Paper (5).
- Crossley, S. A., Skalicky, S., Dascalu, M., McNamara, D. F., & Kyle, K. (2017). Predicting text comprehension, processing, and familiarity in adult readers: New approaches to readability formulas. *Discourse Processes*, 54(5), 14.
- Dale, E., & Chall, J. S. (1949). The concept of readability. *National Council of Teachers of English*, 1-47.
- Dell'Orletta, F., Wieling, M., Cimino, M., Venturi, G., & Montemagni, S. (2014). Assessing the readability of sentences: Which corpora and features? In *Association for Computational Linguistics* (pp. 163-164).
- Dell'Orletta, F., Montemagni, S., & Venturi, G. (2011). READ-IT: Assessing readability of Italian texts with a view to text simplification. In *Association for Computational Linguistics* (pp. 73-83).

Elder, C., & Harding, L. (2008). Language testing and English as an international language. *Australian Review of Applied Linguistics*, 31(3), 341-343.

Klonis, A. [Κλώνης, Α.]. (2019). Κατασκευή λογισμικού επιπέδου γλωσσομάθειας και μέτρησης δυσκολίας τεστ. In Πρακτικά Εργασιών 5ου Διεθνούς Συνεδρίου, για την Προώθηση της Εκπαιδευτικής Καινοτομίας, Λάρισα 11-13 Οκτωβρίου 2019 (pp. 1239-1245). Retrieved 16/12/2020, from shorturl.at/eAFO3.

Lenzner, T. (2014). Are readability formulas valid tools for assessing survey question difficulty? *Sociological Methods & Research*, 43(4), 678-681.

Mackay, N., McCluskey, A., & Mayes, R. (2010). The log handwriting program improved children's writing legibility: A pretest–posttest study. *American Journal of Occupational Therapy*, 64, 30.

Zuffi, S., Brambilla, C., Beretta, G., & Scala, P. (2007). Human computer interaction: Legibility and contrast. In *IEEE*.