

Lexical Density as a Readability Indicator in Translations of Children's Literature

Daniel Nedelcu, "Alexandru Ioan Cuza" University, Iași

Introduction

When it comes to education, a common theme in the press over the past few years has been the poor results Romania gets on the PISA tests (The Programme for International Student Assessment, conducted by the Organisation for Economic Co-operation and Development). The term "functional illiteracy" is often associated with this topic, together with the idea that students do not understand what they read¹. In the same vein, a 2024 Eurostat study revealed that Romanians do not read in general. Less than 30% of Romanians read at least one book in 2023, and less than 7% more than one book². The age-old complaint of "young people do not read anymore" seems to be holding true in this age.

There are many contributing factors to this phenomenon, but an important one is that comprehension is a skill which requires practice. Solving functional illiteracy requires increasing the comprehension level of the reader, which can only be achieved through practice. Reading is as much habit as it is practice –an adult reader is most often the result of a child reader. In a 2023 study, Zare et al. set out to empirically establish this well-known axiom. "Results showed positive links between favourable memories of reading during elementary and high school years and present-day reading habits. Conversely, unfavourable memories during high school were associated with unenthusiastic present-day reading habits" (Zare 262). Therefore, research has established that to ensure an adult population which reads (well), the habit needs to be positively cultivated as early as childhood. Educators, researchers and policymakers need to look into the issue of child reading in order to combat the issue of adult illiteracy (functional or not). Implicitly,

¹ Marina, Georgiana. "Dezastru la testele PISA. România e pe penultimul loc în UE. Elevii români sunt la nivelul de competență..." *Digi24*, 5 Dec. 2023, www.digi24.ro/stiri/actualitate/educatie/dezastru-la-testele-pisa-romania-e-pe-penultimul-loc-in-ue-elevii-romani-sunt-la-nivelul-de-competenta-2-considerat-critic-2604155. Accessed 11 November 2024.

² Eurostat. "Younger People and Women in the EU Read More Books." *Eurostat*, 9 Aug. 2024, ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20240809-2. Accessed 11 November 2024.

the genre of Children's Literature (ChL) becomes an important study object, especially transdisciplinary – literary and cultural studies, linguistics, education, psychology, sociology, policy making.

The converse does not always hold true. An avid child reader does not guarantee that, as adults, they will keep the habit. An interesting phenomenon can be noticed at the border between primary and middle school: gradually, children stop reading. Many reasons come to mind: reduced oversight from teachers as children advance, lack of guidance in choosing the right books, bad experiences with reading, lack of motivation. Furthermore, it seems like reading comprehension is associated with an intrinsic motivation for reading:

Previous studies with students beyond third grade have consistently found positive associations between intrinsic reading motivation and reading comprehension, while extrinsic reading motivation showed negative or no relationships with reading comprehension. These results hold even when controlling for other predictors, such as prior reading competence, verbal ability, and decoding skills. (Stutz et al. 102)

In other words, if a child enjoys what they are reading, they will more easily develop comprehension skills. Comprehension skills also ensure the child has a sense of satisfaction as they continue reading (children easily discard materials which are too hard or pose greater difficulties than the child can carry), which leads to further reading. Stutz et al. claim that “based on the present findings, it may be concluded that significant relations of intrinsic reading motivation with reading amount and reading comprehension are in particular to be expected [...]” (110). Therefore, what a child reads and when they read it can help with intrinsic motivation.

Readability

In order to ensure that children read adequate materials which promote development of reading comprehension and reduce the chance of children giving up on their reading due to difficulties, countries across the world have adopted the concept of “reading level,” wherein books are ranked on a level of difficulty (most often based on readability index) and children and teachers choose appropriate books according to the child's reading level. Most often, a book with a readability level slightly higher than the child's reading level will be chosen, in order to challenge the reader and

stimulate growth, while also ensuring the gap is not wide enough to pose serious reading impediments.

Readability is to the text what comprehension is to the reader. Readability is a trait of the text, while the concept of text comprehension is a cognitive process specific to the reader. “Text readability affects the magnitude of cognitive efforts the pupil must devote to the text processing” (Gavora 4). In other words, if reading is seen as a complex cognitive process, readability is the quality of the text which makes the process easier or harder on the reader. Readability influences comprehension: the easier it is to read a text, the fewer cognitive efforts are needed from the reader. Comprehension, as opposed to readability, is (or should be) the result of the act of reading: the formation of a subjective interpretation and/or understanding of the text by the reader. The uniqueness of the interpretation is given by the variety of factors which influence the process leading to comprehension, such as background knowledge, affective and cognitive processing abilities, and values and attitudes among others. This perspective is important when dealing with ChL as all these elements are influenced by age and bear a stronger impact in the child’s reading experience compared to adults. As a child develops comprehension skills and a reading habit, readability needs to be optimal to the reading level of the reader.

Readability is therefore one of the fundamental factors contributing to comprehension. While the basis of comprehension is the act of reading itself, readability is seen by Boyer as one of the fundamental traits of the text (complemented by subject and materiality) that impact the interaction between reader and text (8). Readability is therefore the totality of linguistic components of the text that facilitate or hinder the process of comprehension.

Taking into account that there are many approaches to readability, there are currently two types of assessments: subjective and objective. Both have their strengths and weaknesses, and neither is superior to the other. However, the best results (and more accurate) will be given by the combination of the two, especially when the subjective assessments are used to control for the objective ones.

Subjective readability assessments suppose the use of scales, questionnaires or interviews to qualitatively and quantitatively gauge reader perception. These may deal with either specific or general readability aspects. Some of the interviews might include factors such as speed of reading or signs of excitement, boredom or confusion. Subjective assessments focus on the readers, their experience and opinion. This poses the issue of identifying when and how non-linguistic aspects

(especially pertaining to the reader) interfere with the assessment of linguistic ones. Subjective readability assessments are highly personal and can be considered as representative for a group only when done on a significant number of its members.

Objective readability assessments imply doing highly specific linguistic measurements. They use a combination of text features such as word length in syllables, length of sentences, word iteration (redundancy), proportion of verbs to substantives, concrete words vs abstract words (Gavora 11). Objective assessments are mathematical expressions of the statistical relationships between various text features and are usually nowadays performed within corpus linguistics.

The discussion on readability measurement cannot avoid the topic of readability formulas or indexes, which are by far the most popular and common way of determining the readability of a text. However, modern linguists have constantly criticized the shallow approach to readability such indexes have, no doubt due to the practicality of and ease of identifying the features used in the calculations (usually word and sentence length, but sometimes the lexis is considered as well, as based on a synthetic gradation list of what word belongs to what grade level, as decided by the educational system or by the researchers themselves, such as in the case of the USA). While such researchers as Bailin and Grafstein support the idea that readability indexes are not accurate by saying that “readability formulas have too many flaws to even offer sound practical advice” (9), Ghoreishi and Aminzadeh explain that psycholinguists and cognitive linguists, some of the staunchest opponents of the shallow-feature readability formulas, have increasingly ceded that surface text features, despite their limitations, continue to provide the more reliable results when assessing readability (247). No doubt, as technology advances and more complex features will be analyzable without too much manual effort on the part of the researcher, so will the readability formulas, hopefully for more languages than just major-culture ones (such as English).

Most of the reading index formulas, however, were conceived in the previous century, some even as far back as the 1940s (Flesch-Kinkaid) or older. Advances in technology have permitted linguists increasing flexibility in usage of readability variables; while some features are easy to identify (such as the total number of iterations of a specific part-of-speech, and redundancy), marking them and inserting them in a calculation program can be cumbersome and time-consuming if done manually. For this reason, some morphosyntactic features that are possibly more relevant to readability than the shallow ones used by most readability indexes are becoming more common in research. Advances in corpus and computational linguistics allow for automated

POS (part of speech) tagging and even parsing entire texts and corpora. Computer applications can even be programmed to teach themselves and start identifying patterns and frameworks on their own (NLP processing, machine learning, and AI). This requires a high degree of programming literacy on the part of the linguist, re-training in linguistics on the part of the computer scientist, or the happy collaboration of the two.

Lexical Density

As mentioned earlier, readability is the totality of linguistic features in a text which promote or hinder comprehension. Traditionally, readability indexes or formulae have taken into account word length. This is, of course, Anglo-centric; the average syllable length in other languages is not as constant as it is in English (see German for example, but also Latin languages, not to mention logographic writing systems such as Chinese). Some languages managed to develop their own readability formulas based on a combination of language features. For example, the Romanian program LEMI (Chitez et al.) considers lexical and semantic features such as the number of archaisms and regionalisms. A feature proposed by Finnish researchers as reflective of ChL readability is the incidence of non-finite constructions (Puurttinen 4).

Lexical density is a feature of the text proposed by researchers such as J. Ure as early as 1971 and M.A.K. Halliday in 1985. Ure's lexical density formula has been traditionally used as a feature of readability, while Halliday's formula as a measure of text complexity, especially in academic text types. Halliday's density deals with the concept of nominalization; a denser Hallidayan text is usually a highly nominalized one, which is a feature of academic and functional text types. Ure's formula, on the other hand, reflects the general information load of a text, how packed it is with information.

Ure's Lexical Density (dU) is expressed through a percentage: how much of the total words are lexical items (To 62). The higher the percentage, the more lexically packed it is. The lower the percentage, the less packed it is. A low dU is important for Children's Literature; it indicates higher percentages of function words, which are conducive to linking meanings and less strain on the cognitive process. Moreover, a low dU indicates the possibility of a higher TTR (type-token ratio) of function words (because they are rather limited in their number), which adds up to the overall readability of the text.

Halliday's 1985 Lexical Density (dH) builds on Ure's earlier one but modifies it to fit within the parameters of his proposed Functional Grammar framework (Halliday 63). His formula is not represented by a percentage; it can start at 0 and can go virtually as high as possible, and it is characterized by the ratio of lexical items to the total number of clauses in a text. While dH is more relevant for more functional text types than for ChL, it can still be used. First of all, nominalization is highly strenuous on the cognitive process, therefore advised against in ChL. A ChL text with a low dH is more acceptable, but it is also expected: strong nominalization has never been a characteristic of ChL. However, it is relevant because it can show predication: it can double as an indicator of verb to other parts of speech ratio (as we can safely infer the majority of clauses in English contain either an active or a non-finite verb); the closer the dH is to one, the higher the verb ratio is. A good verb ratio is one of the most favorable indicators to ChL readability, as it shows the text is characterized by a dynamic narration.

Translating for children

A special mention needs to be made as pertains to Translation of Children's Literature. While the subject of ChL by itself is broad and offers a wide area for study, in small-culture literary traditions, as is the Romanian one, translations form a big slice of the overall ChL available, and are more often than not the child's main reading list; Romanian originals are either not well marketed, or not relevant enough for either the children or their parents (as is the case of most of the government-prescribed Romanian literature syllabus – most books having been written in the previous two centuries.)

A perk of analyzing translated ChL (especially in comparison, where possible) is the fact that, in translation, linguistic norms are much more visible as a result of the battle of cultural transference from one (language) system to another. However good a translator is, the "battle scars" will still be visible to the trained eye and, while not optimal in themselves, fortuitous for this paper. "In fact, the discussion of translated texts is even more fruitful than that of original texts because translational norms expose more clearly constraints imposed on a text that enters the children's system" (Shavit 26). Therefore, even if it might seem more appropriate to analyze native written ChL to discover the specificities of its system, especially in matters of readability, Shavit confirms that analyzing translated ChL might reveal more, exactly because the constraints with

which the translator must work are more visible. The discrepancy between the original and the translation highlights the target culture's expectations about what ChL should look like.

When it comes to the difference between translating ChL and literature aimed at adults, ultimately what needs to be considered is one's perspective on what translation is. If translation is merely a transference of language, text to text, then the source text takes precedence, and the strategies used will mostly be ones that ensure linguistic equivalence; there would therefore be no difference between translating ChL and translating for adults. This, however, is no longer a popular view within the Translation Studies field. As such, the specifics of ChL will reflect the requirements for its translation. The primary trait of ChL is its audience: the children. Therefore, whatever considerations are made for creating ChL will stand true in its translation as well. The Audience takes precedence in this specific situation: "translating for children," as Oittinen puts it (54), rather than "translating children's literature" may be more accurate. The attention shifts from text to audience. While there is markedly a difference between the two concepts, for the purposes of this paper they will be used interchangeably, while maintaining the sense that ChL Translation is translating for children, not a simple transference of text from one language system to another.

Looking at the differences in translating for adults and translating for children, a few small differences can be identified, all of which can be ascribed to either of two areas of ChL-specific problems translators need to be aware of:

1. the specific needs of the target audience and all it entails: cognitive abilities, background knowledge, stage of emotional development, linguistic abilities, expectations (which can be universal, local, generational, individual etc.)
2. expectations of the intermediary groups (parents, teachers, policy makers, editors etc.)

As pertains to the first area, it is even more problematic as various stages of development (cognitive and emotional) form sub-categories of the general audience – children; ChL is never simply created or translated for children, but for children aged 7 and above, for example. Children of a lower age than indicated most likely will not be able to understand the book (neither the language, nor the subject), which might deter them from reading, or worse, forming a reading habit. This delimitation is much clearer in the first stages of a child's development and the first few years after children start learning how to read (Lathey 6).

In general terms, the main difference between translating for children and translating for adults is that there might be a need for more adjustment to the translated ChL text in order to meet expectations in the target culture and make the translated text more acceptable. These adjustments can pertain either to the language used or to the subject of the text. Translators of ChL are in the difficult position of balancing the needs of the target audience (linguistic and cognitive abilities of children) and the norms of literary translation, which were built within a framework developed mainly for the literature of adults. Herein, not surprisingly, comes the dividing formal element between translating for adults, and translating ChL: the recipient.

A child has a much lower tolerance for foreignness and strangeness. A ChL translator must be aware of the limits of this tolerance, and gauge how much foreignness to allow in the text. Otherwise, a child might either simply not understand, missing therefore any benefit of the act of reading (both educational and entertaining), or refuse to continue reading. Foreignness is a big factor in hindering readability as well: the morphology of the words which the child is not used to, the time it takes the child to process the information (and possibly fail), the time it takes the child to look up footnotes or other resources provided or accessible (paratextual elements such as maps, glossaries, guides, or third party resources such as dictionaries, encyclopedias or the internet). The balance between source text orientation and target audience orientation is, therefore, much more relevant in the case of ChL translation, and the translator must ponder and decide based on intrinsic considerations, or extrinsic ones (the intermediary groups).

Looking at the standing of ChL in the polysystem, Shavit goes as far as saying that the reason for which translators allow themselves such liberties is because of its rather peripheral and low position. However, she takes an objective look at it and identifies two situations in which adaptation is the chosen strategy of a translator: when a text is adjusted to be submitted to the social norms of the target culture, and when a text is adjusted to meet the cognitive and linguistic needs of the target audience (Shavit 172).

Oittinen, herself a ChL translator and author in Finland, takes a diametrically opposed position: she affirms that adaptation comes from “love for children and their literature” (Oittinen 58). Through adaptation, translators speak the children’s language and keep their literature (and interest in literature) alive.

Thomson-Wohlgemuth, citing Emer O’Sullivan’s ideas, formulates questions that translators should ask themselves when they start doing a ChL translation or when choosing

whether to domesticate or foreignize a text or a situation: “How much foreignness can a young reader take? Which cultural peculiarities need to be adapted, and which may stay unmodified? How many explanations does a child need in order to be able to cope with foreign issues” (Thomson-Wohlgemuth 40).

A ChL translator needs to be highly intentional in the lexis and syntax used in translation. At various ages, children have a very specific level of linguistic development: a very specific list of words they know (which grows less specific as the child is older and their experiences are diversified) and their abilities to comprehend complex grammar and syntax are lower as they are younger and grow as they are older.

Case study: Lewis Carrol’s *Alice’s Adventures in Wonderland*

The corpus extracted is meant to highlight practices and naturally occurring linguistic situations of ChL translations in Romanian and their readability acceptability for their target audience age group. For this paper, *Alice’s Adventures in Wonderland* by Lewis Carrol was chosen as the corpus for analysis since it is one of the most translated children’s books in Romanian, as well as one of the most representative and recognizable books for ChL, following its many multi-media adaptations since its publishing. Moreover, the target audience for the Romanian translations, based on the book’s appearance on school reading lists and publishers’ recommendations, is the 10-12 age group (4th-5th grade the Romanian school system). This is relevant as it coincides with the passage from primary school to lower-secondary school, which is usually when children start losing their reading habit.

The body of the corpus is comprised of excerpts from the original 1865 *Alice’s Adventures in Wonderland* edition (EN1) as found on the Project Gutenberg website (Macmillan & co., London, 1865) and their counterparts from three different Romanian translations: Bodgan Popescu, 1971 (Ro71); Aura Brais, 1999 (Ro99); Prestige, 2018 (Ro18). For the last translation, the publishing house chose not to disclose the name of the translator, therefore, the name of the publisher will be used to refer to it.

The goal was for the excerpts to be representative of the whole book from which they were extracted. Therefore, approximatively 5000 words (5335 tokens), representing 20% of the entire original *Alice’s Adventures in Wonderland* were analyzed, and their corresponding excerpts in the

three translations: 4635 tokens (Ro71), 5207 (Ro99), 5082 (Ro18). In total, the analyzed corpus consists of 20,259 tokens.

The excerpts themselves were chosen based on the principle of representativity as well. The criteria for choosing the excerpts were

- (1) that they are from different parts of the book,
- (2) that they cover the main modes of writing (dialogue, narration, description),
- (3) that they include some of the most common translational problems the book poses (humour, poetry, names), and, finally,
- (4) that they are iconic scenes and easily recognizable.

The chosen excerpts were the entirety of chapter one (2166 tokens), the Mad Tea Party scene in chapter seven (1746 tokens), and most of chapter twelve (1423), including the trial of the Knave and Alice's return from Wonderland. Chapter one is mostly narration and description, chapter seven is almost entirely dialogue riddled with nonsense, and chapter twelve consists of dialogue and description, containing puns, nonsense, and poetry.

The Romanian (Ro71, Ro99, Ro18) corpus was chosen based on the considerable temporal distance between the three translations (28 years between the first two, and 19 between the last two, covering a total span of almost half a century). The distance is even more relevant as each of the translations were produced in very different eras of the Romanian consciousness: communism, turn-of-the-millennium post-communism, the European Union era. Moreover, as the target audience is highly specific and important to the end product when it comes to ChL, the temporal rift between the three translations represents three considerably different audiences: the first generations born in communism, for the 1971 translation, the millennials for the 1999 translation, which are also the first generation to either be raised or born after communism, and GenZ for the 2018 translations, which are entirely digital natives and whose book consumption habits are by far the most different from any other preceding generation precisely because of their relationship with the digital.

For Romanian, there are unfortunately no (available) studies to indicate what the average lexical density is for a specific target audience or overall in literature. Such studies would be welcome, especially if they would compare the average lexical density in translated literature and the average lexical density in original Romanian literature. A diachronic perspective would

perhaps be even more interesting: how has lexical density changed over time, and whether writers/translators are influencing each other within the Romanian literary system.

In regards to Halliday's density, dH, the way it was calculated for the purposes of research makes it a good indicator of predication in Romanian; as dH is calculated by the total lexical items divided by the total clauses, and clauses were counted only as they contained a verb in a finite form (Romanian syntax does not confer the role of clause to NCs), a lower dH will indicate a higher dynamism of narration, and is therefore favorable.

On the other hand, Ure's density, dU, is calculated as the ratio of lexical items in a text, that is how many lexical items there are compared to the total tokens of the text. What dU indicates is the airiness of language in relation to the information load. A lower dU indicates a higher number of grammatical items (words that have a highly specific and limited function in the discourse). As these grammatical items are limited in number, a lower dU will also indicate a high redundancy among function words. This is positive when it comes to ChL as function words also aid in the clarity of text, ensuring a clear relation between the various items of the sentence and the meaning; a high percentage of grammatical words is conducive to unambiguous linking. The redundancy will also factor into easing the stress on the cognitive process, as repeating the words creates and strengthens neural pathways specific to those items and their collocations or semantic clusters. As in this case they are function words, the strengthening of their neural pathways is all the more beneficial as it also results in faster and clearer semantic and grammatical linkage, which aids in the prediction aspect of the reading cognitive process.

Looking at Table 1, we can see that the Romanian corpora are fairly similar in matters of lexical density. For dH, all three fall in the bracket of 2.5-3 dH. The interpretation of these numbers is that the average clause in Romanian contains between 2.5 and 3 lexical words. As they are all close to the lowest limit of density (which is 1, representative for texts in which the only lexical items are mostly verbs), it is safe to assume all three are characterized by a highly dynamic narration and low nominalization. The verb and noun ratios confirm these findings.

Corpus	Ch	dH	dU
En	1	2.33	41.46
En	7	2.27	49.08
En	12	2.58	44.06
En	avg	2.39	44.87
Ro71	1	2.63	47.51
Ro71	7	2.32	55.44
Ro71	12	2.73	50.87
Ro71	avg	2.56	51.27
Ro99	1	2.85	47.49
Ro99	7	2.46	55.17
Ro99	12	3.17	54.08
Ro99	avg	2.83	52.25
Ro18	1	2.70	48.65
Ro18	7	2.34	55.14
Ro18	12	3.01	52.01
Ro18	avg	2.69	51.93

Table 1: Halliday and Ure lexical density indexes as divided by chapter and overall corpora

The lowest dH is encountered in chapter 7 of the Ro71 corpus at 2.32 dH, while the highest is 3.17 dH in chapter 12 of Ro99 corpus. Lowest dH is encountered respectively in all three corpora in chapter seven (2.32 dH Ro71; 2.46 dH Ro99; 2.34 dH Ro18), which is explained by the fact that the main mode of exposition in chapter 7 is dialogue in a high proportion. Spoken language is automatically lower in density than written language, and the same is true about dialogue vs narration. The inference here is that all three translators translated with a higher degree of speakability in chapter 7 than they did in the other two, which is consistent with the style of the original En corpus where ch1 and chapter 12 are 2.6%, respectively 13.6% denser than chapter 7. The higher overall dH in chapter 12 in the Romanian corpora is explained by the stylistic choices in the source text, where there are a number of long enumerations and tightly packed descriptions. The translators of the 2018 and 1999 versions chose to preserve these parts and used mostly equivalence in

translation, while the translator of the 1971 version chose to break some of them up by introducing verbs or changing word classes in some situations, from noun to verb. This can be seen in the vastly different dH for chapter 12 between the Ro71 and the more recent corpora (16.1% less dense than Ro99 and 10.2% less dense than Ro18). However, one other reason for the lower density in the Ro71 corpus are the frequent hidden abbreviations (or abridgements) done by the translator; this can be observed to be an overall strategy as the total number of words of the corpus is 12% lower than the Ro99 corpus and 10.6% lower than the Ro18 corpus.

On average, the Ro71 corpus seems to be the least dense one, at 2.56 dH, while Ro99 seems to be the densest one at 2.83 dH on average, with Ro18 being the middle ground at 2.69 dH. This would indicate a more dynamic style of narration on the part of the Ro71 corpus, which is confirmed by the verb ratios, which are the highest among the three. The verb to noun ratio and the verb to total lexical tokens ratio is even higher than the English corpus, making it a more dynamic translation than the source text. This can still be explained to some degree by the hidden abridgements made by the translator. But when comparing dH and dU between Ro71 and the

English original, while the verb ratios are higher in the Ro71 by 4.7% (verb to noun ratio), respectively 1.8% (verb to lexical items ratio) and 16.2% (verb to total tokens), the dH is 7% lower in the original. Furthermore, the dU is 14% lower in the original, both of which indicate a higher information load in the text, and slightly higher nominalization in the translated text of Ro71 due to the higher dU (which is explained by the higher adjective ratio, the highest of all 4 corpora).

Nonetheless, despite this peculiarity, the Ro71 can be perceived to be the more dynamic text of the three translations (or the 4 corpora in general), as well as the airiest and least packed. However, the dU does not necessarily provide a relevant difference. While the differences in dH go up to 10.3% (higher dH in Ro99 by comparison to Ro71), the differences in dU between any of the three corpora are around 1%. This leads to the idea that the ratio between function words and lexical words is around 50% for Romanian, but more relevant general characterization should be performed on larger corpora, which should include original Romanian texts as well.

However, subsequent subjective measurements with 14 children aged 10-12 invalidated the objective result, as all participants chose the Ro71 translation to be the most complicated one to read, as well as the least enjoyable. The underlying reason for this result was revealed in one of the subjective tasks which asked the children to indicate words or phrases they do not know, and Ro71 has the most of them.

A similar result was rendered by the LEMI tool (Chitez et al.), an online readability analysis tool specializing in literary texts for children. While the analysis of fragments from the Ro71 corpus consistently showed grammatical features which should ensure a lower readability score compared to the other two corpora, its lexical features (i.e. number of regionalisms and archaisms, as per the LEMI algorithm) caused its LEMI readability index to be higher than the other two (Ro71 marked as grade 5, level 3 complexity; Ro99 and Ro18 marked as grade 5, level 2 complexity).

While this may seem that density is not necessarily a good indicator for readability in the case of Children's Literature (and its translations), the subjective measurements indicate a different possible conclusion, which needs further testing: optimized grammatical and syntactical features (including density) which improve readability are a pre-requisite of an enjoyable text. The subjective measurements asked the children to compare the same fragment from all three translations, and a fourth additional one which was translated intentionally to optimize grammatical and syntactical features to ensure the easiest possible reading experience, while not

compromising on the style and content of the text. Unanimously, the fourth fragment was chosen as the easiest to read, as well as the most enjoyable. Furthermore, besides the Ro71 fragment, the subjects did not indicate any lexical issues, which means that the only difference between the Ro99, Ro18 and the fourth fragment was of the morpho-syntactical order. As such, since in the absence of lexical issues, children are affected by the grammatical and syntactic features of the text (of which density is one), it can only be concluded that these features are expected to be optimized for an appropriate readability level. Density by itself, however, is not a good predictor of readability in Romanian, especially in the absence of a lexical component to the algorithm.

Conclusions

Lexical density refers to the information load or how tightly information is packed in the text. A higher lexical density indicates a highly informative text, most likely characterized by strong nominalization especially with Halliday's lexical density formula). In Children's Literature we look for (and expect) a lower lexical density, as a tightly packed text lexically requires increased cognitive effort on the part of the reader and is also characterized by ambiguity regarding semantic relations between elements. Meaning is not immediately obvious, while the high load of information strains cognitive capacity. All these are not favorable for children, as increased effort and frustration resulted from lack of understanding lead to a lower motivation for the child to finish the book.

Based on a corpus analysis of three Romanian translations of *Alice in Wonderland* (Ro71, Ro99, Ro18), the density index of each indicated that the Ro71 translation is favorable over the others due to its lower density (according to both Ure and Halliday's formulas). Subjective measurements, however, invalidated this result, as children belonging to the target audience chose the Ro71 text as the most difficult to read. At the same time, subjective measurements indicated that, while density itself is not a primary determinant of readability, it is nonetheless expected that a text should have an acceptable density level as a basis for good readability.

The limitations of the research are the low number of participants in the questionnaire, the lack of Romanian standards for the text features discussed, lack of other subjective measurements to cross-reference (interviews, cloze tests), and the scarcity of studies done in this field.

Research into the readability of ChL and its translation in Romanian is important in the context of a literacy crisis, of a growing percentage of children who do not read, and of a growing gap between socio-economic groups in matters of literacy. Education experts, academics, and policy makers need to invest in studies on literacy and provide tools and policies for improving the relationship between children and reading. Readability plays an important role as seen above and should be investigated in order to describe realities and to propose solutions and directions. Realities that need determined are the levels of readability in Romanian, the children's ability to read in comparison to the level of the books they are assigned, the Romanian educational system's expectations, and the European Union's expectations in comparison. Some solutions and directions that need to be formulated are lists of books suitable for each developmental stage/age, precise standards for translation in regards to readability, lists of words expected for a child to know at a specific age/developmental stage, as well as studies and tests for measuring Romanian children's reading level.

Works Cited

- BAILIN, Alan and GRAFSTEIN, Ann. *Readability: text and context*. Palgrave Macmillan, 2016.
- BOYER, Jean-Yves. "La lisibilité." *Revue française de pédagogie*, vol. 99, 1992, pp. 5-14.
https://www.persee.fr/doc/rfp_0556-7807_1992_num_99_1_1322. Accessed 11 November 2024.
- CARROL, Lewis. *Alice în țara minunilor*. Edited by Nicoleta Popescu. Editura Prestige, 2018.
- CARROL, Lewis. *Alice în țara minunilor*. Translated by Aura Brais, Editura Coresi, 1999.
- CARROL, Lewis. *Alice în țara minunilor*. Translated by Bogdan Popescu, Editura Didactică și Pedagogică, 1971.
- CARROL, Lewis. *Alice's Adventures in Wonderland*. Produced by Arthur DiBianca and David Widger, 2008. Project Gutenberg, www.gutenberg.org/files/11/11-h/11-h.htm . Accessed 11 November 2024.
- CHITEZ, M., Strilețchi, C., Rogobete, R., Oravițan, A., Csűrös, K., Andercău, F., Prohap, D. R. & Strilețchi, V. (2023). *LEMI – Platformă de evaluare a complexității lingvistice a literaturii pentru copii în limba română*. Universitatea de Vest din Timișoara. Available at: <http://www.lemi.ro/>.

- GAVORA, Peter. "Text Comprehension and Text Readability: Findings on Lower Secondary School Pupils in Slovakia". *Forums Dydaktyczne*, vol. 9-10, 2012, pp. 9-21.
<https://repozytorium.ukw.edu.pl/bitstream/handle/item/1770/Peter%20Gavora%20Text%20Comprehension%20and%20Text%20Readability%20Findings%20on%20Lower%20Secondary%20School%20Pupils%20in%20Slovakia.pdf?isAllowed=y&sequence=1>.
Accessed 11 November 2024.
- HALLIDAY, M.A.K. *Spoken and Written Language*. Deakin University, 1985.
- LATHEY, Gillian. *Translating Children's Literature*. Routledge, 2016.
- OITTINEN, Riitta. *Translating for Children*. Garland Publishing Inc., 2000.
- PUURTINEN, Tiina. "Syntax, Readability and Ideology in Children's Literature". *Meta*, vol 43 issue 4, 1998, pp. 524-533. <https://www.erudit.org/en/journals/meta/1998-v43-n4-meta169/003879ar.pdf>. Accessed 11 November 2024.
- SHAVIT, Zohar. "Translation of Children's Literature." *The Translation of Children's Literature: A reader*. Edited by Gillian Lathey. Multilingual Matters, 2006, pp. 25-39.
- STUTZ, Franziska, SCHAFFNER, Ellen, SCHIEFELE, Ulrich. "Relations among reading motivation, reading amount, and reading comprehension in the early elementary grades." *Learning and Individual Differences*, vol. 45, 2016, pp. 101–113.
<https://doi.org/10.1016/j.lindif.2015.11.022>. Accessed 11 November 2024.
- THOMSON-WOHLGEMUTH, Gabriele. *Children's Literature and its Translation. An Overview*. Unpublished. MA Thesis, University of Surrey, 1998.
https://translationjournal.net/images/e-Books/PDF_Files/Childrens%20Literature%20and%20its%20Translation.pdf. Accessed 11 November 2024.
- TO, Vinh and Fan, Si. "Lexical Density and Readability: A Case Study of English Textbooks". *The Internet Journal of Language, Culture and Society*, Issue 37, 2013, pp. 61-71.
<https://aaref.com.au/wp-content/uploads/2018/05/37-07.pdf>. Accessed 11 November 2024.
- ZARE, Manzar, et al. "The roots of reading for pleasure: Recollections of reading and current habits." *Literacy*, vol. 57, no. 3, Feb. 2023, pp. 262–74. <https://doi.org/10.1111/lit.12315>. Accessed 11 November 2024.