
A Formulaic Approach to Propositional Density and Readability

Meisam Ziafar¹ and Ehsan Namaziandost^{2*}

¹ Assistant Professor, Department of English Language Teaching, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran.

² Ph.D. Candidate in TEFL, Department of English, Faculty of Humanities, Shahrekord Branch, Islamic Azad University, Shahrekord, Iran.

*Corresponding author email id: e.namazi75@yahoo.com

Date of publication (dd/mm/yyyy): 05/01/2020

Abstract – There are two general approaches to readability: classic and cognitive. There are more than 50 classic procedures for predicting text difficulty, which are usually referred to as readability formulas. Propositional density has been proposed as one criterion for determining readability. Some criticisms have been proposed against taking propositional density as a criterion for readability and such criticisms would be best refuted through taking formulaic expressions as idea units to be taken into account when determining readability. After going through a library research, the authors of this study came arrived at the conclusion that in decomposing utterances into their constituents, taking multiword expressions (formulaic expressions) as propositional units is quite telling and beneficial.

Keywords – Readability, Multiword Expressions, Propositional Density, Formulaic Expressions, Text Difficulty.

I. INTRODUCTION

Chall (1996) further maintains that about the middle of the 1970s, dissatisfaction with classic readability grew quite strong, despite the fact that the readability formulas continued to be used widely. The major dissatisfaction with the classic readability formulas put forward by the critics was that they measured mere surface factors, not real sources of difficulty. Some critics accepted that the classic formulas had practical validity and value, but were essentially a-theoretical. What was needed instead, it was claimed, were schemes based on cognitive theory that would explain text difficulty, and could be used more effectively for estimating difficulty and for writing and editing readable text.

Chall asserts that the two broad and important factors affecting readability proposed by Kintsch and Vipond (1977, cited in Chall, 1996) are the propositional density of a text and the number of new concepts per proposition. More, specifically, the factors included in their analysis were: density of propositions, and the number of different arguments, coherent parts, inferences required to connect a text base, long-term memory searches and reinstatements of propositions into short term memory, and reorganizations required to arrive at the best organized text base.

There are some criticisms against traditional readability formulas. According to Crossley, Greenfield and MacNamara (2008) although traditional readability formulas such as Flesch reading ease and Flesch-Kincaid grade level have been accepted by the educational community, they have been widely criticized by both first language (L1) and L2 researchers for their inability to take into consideration deeper levels of text processing, cohesion, syntactic complexity, rhetorical organization, and propositional density. They further believe that in due course, a readability measure would need to be framed that takes appropriate account of the role of working memory and the constraints it imposes in terms of propositional density and complexity. Crossley, Alan and MacNamara (2011) also believe that traditional readability formulas have also been faulted in the production of L2 texts because they do not account for reader characteristics or text-based factors such as syntactic complexity,

rhetorical organization, and propositional density.

Crossley, Louwse and McNamara (2009) assert that discontent in L2 research with traditional readability formulas led to researchers eschewing the formulas and using human evaluations of text difficulty to determine text readability for simplified and authentic texts. Many L2 reading specialists resorted to conducting large-scale, extensive qualitative studies of simplified reading texts in an attempt to assess their readability. These assessments depended on the evaluative responses of reading specialists and L2 teachers in either placing texts within hierarchies of difficulty or assessing texts for specific linguistic factors (e.g., vocabulary, grammar, and density of information). While these studies were valuable, they tended to be based on the subjective assessment of the graders and the textual information provided by the publishers. Moreover, they were time consuming and constrained in the number of lexical, grammatical, and discourse features that they could assess; thus, limiting the use of the measures for large corpora of L2 texts.

Crossley, Dufty, McCarthy, & McNamara propose Coh-Metrix, which they believe offers the prospect of addressing the limitations of conventional readability measures by providing detailed analyses of language by integrating lexicons, pattern classifiers, part-of-speech taggers, syntactic parsers, shallow semantic interpreters, and other components that have been developed in the field of computational linguistics. In reference to cohesion indices, Coh-Metrix also analyzes co-referential cohesion, causal cohesion, density of connectives, Latent Semantic Analysis metrics, and syntactic complexity. Since Coh-Metrix considers text-base processing and cohesion, it is well suited to address many of the criticisms of traditional readability formulas.

Liu, Su, Lai, Yand, and Hsieh (2009) provide some criticisms against each of the traditional readability formulas. They assert that the criticisms of the traditional readability formulas by the various authors have a lot in common. They all urge adoption of language-oriented criteria based on independent evidence and a closer re-examination of the genuine relationship between the variables and the texts. It is our belief that this can only be done if we take account of the deeper levels of text processing. Reading is a multidimensional process; our pilot study aims to examine how a reader interacts with a text at the lexical level. We propose that the hierarchical status of a lexical item in our mental lexicon is a possible factor that affects lexical comprehensibility. We further suggest that there is a basic level in the lexical hierarchy which is the easiest to comprehend and serves as a meaningful indicator of text readability.

Carrell (1987) asserts that as the foregoing studies demonstrate, the real problem with readability formulas lie not only in the textual features they ignore (syntactic complexity, textual cohesion, propositional density, rhetorical structure) but in the fact that they ignore the reader and reader variables such as background knowledge. They ignore the interactive nature of the reading process- the interaction of the reader with the text. They ignore the fit between the various features of text, including its content domain, its rhetorical organization, its perspective on the topic on the one hand, and on the other hand, the various schemata of the reader- the reader's schemata for content and rhetorical organization, the reader's background knowledge, experiences and beliefs.

Carrell (1987) believes that what is needed is a clearer theoretical approach to readability, one which takes a broader range of reader as well as text variables into consideration. Valid measures of a text's comprehensibility, particularly to a second or foreign language reader, require consideration of a number of factors. Those with responsibility for selecting reading materials which will appropriately match ESL readers' abilities should consider the following:

1. Consider whether the reading is to be done extensively, outside the classroom and without the support of a teacher, or intensively, inside the classroom with support of a teacher.
2. Take into account other recognized contributors to comprehension, especially higher levels of interest, motivation or prior knowledge. Particularly interesting texts and appropriate levels of prior knowledge can stimulate motivation and keep challenging material from being frustrating.
3. Bear in mind that readability formula scores, deriving from counts of style difficulty, become poorer predictors of difficulty for older, more mature readers, where content weighs more heavily.
4. Factor into readability judgments, consideration of textual phenomena at the level of discourse, for example cohesiveness, coherence, the flow of topics and comments, and propositional density-factors which have been shown to play a part in comprehensibility.
5. Factor into readability judgment, consideration of syntactic and lexical choices other than those which affect length- e.g. quasi-technical vocabulary, or non-technical vocabulary used in technical ways.
6. Consider rhetorical structure of texts and the transparency of the signaling of that rhetorical structure, both of which have been shown to affect reading comprehension.
7. Pay special attention to any specialized background knowledge which the text may presume of the reader, and whether the reader has such background knowledge, or can and will be helped in some way to relate existing background knowledge to that presumed by or presented in the text.
8. Consider the trade-off between effort to be expected and value received by the reader in terms of information, entertainment, or whether the purpose for reading.

What seems a more acceptable criterion which determines the reading time is the number of propositions. According to Embretson and Wetzel (1987), Kintsch and Keenan (1973) demonstrated that the number of propositions in a sentence rather than the number of words, determines reading time. This finding validates the use of propositions rather than separate word counts for assessing text processing. If the number of words is held constant, texts with relatively more propositions should be more difficult. Thus propositional density (the number of propositions divided by the number of words) should be related to item difficulty. Mitchell's (1983, cited in Embretson & Wetzel, 1987) results, however, were inconsistent with respect to the impact of propositional density on item difficulty. This inconsistency may be accounted for by separating propositions into the convenient typing. Propositional density was therefore scored separately for predicate, modifier, and connective propositions.

Kinstch, Kozminsky, Sterby, McKoon, and Keenan (1975) found that reading times were longer and recall was less for texts with many different arguments than for texts with fewer arguments. Propositions containing new arguments require an additional processing step on the part of the reader. Argument density is the number of unique arguments divided by the number of words, and the expectation was that higher density would be associated with less difficult items. Another finding from propositional analyses is that propositions are not equally difficult to remember, with superordinate propositions being recalled better than propositions that were structurally subordinate.

Turner and Greene (1978) assert that within the Kintsch's propositional theory the meaning of a text is assumed to be represented by a list of connected propositions. This proposition list is called a text base. Propositions consist

of abstract word concepts. These are the entries in the lexicon of semantic memory, and may be represented by one or more words in the surface structure. Propositions have been classified into three distinct classes, based on the type of relation they contain: predication, modification, and connection. Predicate propositions express ideas of action or states. Modifier propositions change a concept by restricting it or limiting it by means of another concept. Connective propositions relate propositions or facts in the text to each other. These propositions may be left unexplained in the text and must then be inferred. Whole propositions are often embedded as arguments in other propositions.

1.2 *Propositional Density and Readability*

Newbold and Gillam (2010) assert that when a significant amount of information is conveyed in a relatively small amount of text, the reader can become confused. We identify this problem as ‘Propositional Density’. According to DeFrancesco and Perkins (2012), propositional density analysis involves the examination of text complexity. Readability formulas are not concerned with this and propositional analysis also embraces text understanding and retention. Although long collocations form semantic units that reduce conceptual complexity, problems occur when numerous semantic units are described within a short space of each other causing the reader to make numerous inferences. The number of ideas expressed in the text contributes to the work required of the reader to interpret the text correctly. Propositional density may be measurable by examining the quantity of objects within short distances of each other. These objects can be labeled with single nouns or multi-word expressions. By measuring the number of unique semantic units, we can approximate the workload required for processing or interpreting the text correctly.

Some criticisms can be proposed against taking propositional density as a criterion for readability and such criticisms would be best refuted through taking formulaic expressions as idea units to be taken into account when determining readability.

1. The most important justification so far proposed for a propositional theory is the intuition which says that it works as a tool in experimental work and has some psychological validity.
2. The concept of proposition is vaguely defined and results in multiple interpretations.
3. It complicates the issue due to bringing up issues such as subjectivity and fuzziness of typical members. Turner and Greene (1978) assert that nominal propositions must be inferred based on ambiguous probabilistic information.
4. Cohesiveness and cohesion advocated by a propositional theory is highly bound to a linguistic and structural understanding which deprives it from taking meaning (ideas) as the main constituents connected together through propositions.
5. Through restricting itself with grammatical units such as subordinate clause and main clause, a propositionalization of texts proves to be highly bound to decomposing formal aspects of language rather than meanings and ideas. Propositions seem to be inadequate for mapping units of meaning to linguistic forms.
6. Rather than easily taking formulaic expressions as propositional units, a propositional theory chooses to take the less convenient route of mapping well defined meaning boundaries to vague and ambiguous linguistic constituents.

7. Turner and Greene (1978) assert that such system is involved with the semantic content of texts and is devoid of pragmatic considerations. The claim is that even in the case of semantic content Propositional theory has not been successful due to being extremely inclined towards formal aspects of language rather than semantic contents.

II. CONCLUDING REMARKS

The authors of this study gone through a library research and after reviewing previous studies they came to a number of specific conclusions. In decomposing utterances into their constituents, multiword expressions (formulaic expressions) can be taken as propositional units. Multiword expressions must be taken into account in determining readability not only as linguistic pieces which affect the rate of propositional density, but also as pieces influencing the quality of propositions which have been neglected in definitions provided so far for propositional density. No matter how many propositions may be present within a stretch of 100 words, what counts is the quality of such propositions according to their formulaicity and their collocative qualities. The meaning of a text seems to lie within its connected multi-word expressions rather than vaguely defined concept of proposition which is asserted to represent idea units. A new readability formula could be defined based on the number of multi-word expressions within a certain number of words. The claim may be that the more the number of formulaic units the better one's reading. The main advantage of such formula would be the fact that it has psychological justifications due to the presence of research which proves the facilitative role of formulas in receptive and productive skills. This position provides a readability formula with a sound theory and takes readability as an interactive phenomenon through which both the reader and the text play a role.

REFERENCES

- [1] Carrell, P.L. (1987). Readability in ESL. *Reading in a Foreign Language*, 4(1), 21-40.
- [2] Chall, J.S. (1999). Varying approaches to readability measurement. *Revue Québécoise de Linguistique* (Quebec Journal of Linguistics), 25(1), 23-40.
- [3] Crossley, S.A. Greenfield, J. & MacNamara, D.S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, 42(3), 475-493.
- [4] Crossley, S.A., Alan, D.B., & MacNamara, D.S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language*, 23(1), 84-101.
- [5] Crossley, S.A., Duffy, D.F., McCarthy, P.M., & MacNamara, D.S. (2007). Toward a new readability: A mixed model approach. In D.S. MacNamara and G. Trafton (Eds.), *Proceedings of the 29th annual conference of the Cognitive Science Society* (pp. 197- 202). Austin, TX: Cognitive Science Society.
- [6] Crossley, S.A., Louwrese, M., & MacNamara, D.S. (2009). Identifying linguistic cues that distinguish text types: A comparison of first and second language speakers. *Language Research*, 42, 361-381.
- [7] DeFrancesco, C. and Perkins, K. (2012). An analysis of the proposition density, sentence and clause types, and nonfinite verbal usage in two college textbooks. In M.S. Plakhotnik, S.M. Nielsen, & D.M. Pane (Eds.), *Proceedings of the 11th Annual College of Education & GSN Research Conference* (pp. 20-25). Miami: Florida International University.
- [8] Embretson, S.E., & Wetzel, C.D. (1987). Competent latent trait models for paragraph comprehension tests. *Applied Psychological Measurement*, 11(2), 175-193.
- [9] Kinstch, W., Kozminsky, E., Sterby, W.J., McKoon, G., & Keenan, J.M. (1975). Comprehension and recall of text as a function of content variable. *Journal of Verbal Learning and Verbal Behavior*, 14, 196-214.
- [10] Y., Su, C.C., Lai, Y.D., & Hsieh, S.K. (2009). Assessing text readability using hierarchical lexical relations retrieved from Word Net. *Computational Linguistics and Chinese Language Processing*, 14(1), 45-84.
- [11] Newbold, N., & Gillam, L. (2010). The linguistics of reliability: The next step for word processing. *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing, USA*, 65-72.
- [12] Turner, A., & Greene, E. (1978). *The construction and use of a propositional text base*. Boulder, CO: University of Colorado.

AUTHOR'S PROFILE



First Author

Meisam Ziafar received his PhD degree in English Language Teaching from Islamic Azad University, Science and Research Branch, Tehran, Iran. He is an assistant professor at the Department of English Language Teaching, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran. His research interests are contrastive studies, second language acquisition, pragmatic competence, and foreign language teaching methodologies. email id: meisam-ziafar@iauhavaz.ac.ir



Second Author

Ehsan Namaziandost was born in Shiraz Province of Iran in 1985. He holds an MA degree in TEFL from Islamic Azad University of Ahvaz. Now, he is Ph.D. candidate of TEFL at Islamic Azad University, Shahrekord, Iran. His main interests of research are second language acquisition, EFL teaching and testing, teaching language skills, and language learning strategies. His research papers and articles have been published by different international journals.
email id: e.namazi75@yahoo.com