A Syntactic Analysis of Cambodian News Discourse on COVID-19 Outbreaks: Sentence Lengths and Structures as Predictors of Readability

Bunlot Khoy, Sotheara Suon & Bophan Khan,
Faculty of Social sciences and Humanities, Royal University of Phnom Penh, Cambodia; Faculty of Social sciences and Humanities, Royal University of Phnom Penh, Cambodia & Institute of Foreign Languages, Royal University of Phnom Penh, Cambodia

Abstract

The purpose of the study is to conduct a syntactic analysis of news discourse published online in four popular Khmer newspapers in Cambodia. A sample of 60 articles on COVID-19 outbreaks published from 23 to 29 January 2020 were analysed qualitatively and quantitatively, and with the PAN Khmer Line Breaking Program used to separate Khmer words for word count the way English words are counted as Khmer writing conventions do not require a space in between words but phrases. The researchers used descriptive and inferential statistics to calculate and compare average lengths of the articles, paragraphs, and sentences. The study compared its results to subject-specific academic discourse for Grade 12 and undergraduate studies, and found no statistically significant difference in average sentence lengths between academic texts and the target articles, with sentences of the newspaper articles averaging 32.25 words, compared to 34.3 for the academic texts. Complex sentences were most commonly used, accounting for 44% of a total of 442 sentences in the study's dataset. Based on the findings, Khmer journalistic writers are recommended to reduce complexity of clauses and sentences in their written news pieces, observing the standard readability of under 20 words per sentence.
Keywords: COVID-19, News Discourse, Readability, Syntactic Analysis, Sentence Structure

1. Introduction

This section presents the background of the study, context of media publication in Cambodia, purpose of the study, and research questions. Section 2 reviews related literature and presents the analysis framework of the study, while Section 3 presents findings of the study for both research questions. The last section of the article compares the study's findings to those of the past studies, and provides recommendations for professional training of Khmer journalistic writers and future research.

Newspapers are a staple item that most people come in contact with every day. For some, reading newspapers is a necessity, perhaps in the form of a morning ritual, in order to keep up with what is happening in the world, while it is an ideal activity for relaxation for others (Järvbäck Hillbom, 2009). Plus, newspapers are believed to have more readers than any other type of written text because it consists of articles of an enormously miscellaneous nature and has therefore become a significant source of information across the world. Various information that is aimed at addressing the needs of its perceived readership is provided through news and sports analyses, feature articles and advertisements (Grossberg et al., 1998). In this way, content and language use in newspaper writings form an interesting relationship between the providers and receivers of information. In terms of language use, sentence complexity is found to have direct impact on readability of a discourse (Biber et al., 1999).

In Cambodia, in line with the global trend, traditional print news media are gradually replaced by online news outlets. Okonofua (2010) captures this trend, crediting innovations in Information and Communication Technologies (ICT) as an impetus leading to multiplicity in newspaper publishing, including online versions, with an increased ability to connect and

interact with readers. Many daily newspapers in Cambodia now also have online versions in addition to their traditional print counterpart, while at least half of the local newspapers publish only online. Bittner (2003) stated that the first newspaper, Columbus Dispatch, to cater to Internet users went online on July 1, 1980. In addition to the conventional online space, newspapers nowadays are commonly featured in various social media platforms, appealing to social media users (Nkemdilim, 2015), making news content available within scrolls and clicks.

This study investigates language use in Khmer journalistic writing, aiming to provide a description of features of Khmer professional newspaper writing style to inform implications about Cambodia's newspaper readability, comprehensibility and other relevant syntactic aspects. This study begins with the premise that the writing style or linguistic choices play a vital role in the readability and comprehensibility of newspapers (DuBay, 2004). Writing style here implies the lexical and syntactic elements used to construct newspapers reports. The study begins with a proposition that the language of Khmer newspapers is complex lexically and syntactically, and that this could negatively impact the readability and comprehensibility of the newspapers among their target consumers. Complexity in this study refers to the occurrence of syntactic categories and constituency in a text in a manner that could compromise knowledge transfer and in doing so cause some readers understanding complications (McNamara, Crossley, & McCarthy, 2010). In the context of Cambodia, not just as a developing country but also as an emerging democracy, problems related to the ability of the media to communicate with its readers with clear language use are enormously significant. Therefore, critical questions that cannot be ignored and which this study endeavors to address are as follows:

(1) What is the average length of articles, paragraphs, and sentences published in four popular online newspapers in Khmer, in comparison to academic texts for Grade 12 and undergraduate students?
(2) What types of sentences are most commonly used in the articles published in the four newspapers?

2. Literature review

News discourse is commonly subject to Critical Discourse Analysis (Fairclough & Wodak, 1997), with a good proportion dedicated to the study of headlines (Hassan, 2018) and focusing on relationships between language choices in the news pieces and ideological implications for the readers (Chen, 2017; Hassan, 2018; Pang & Wu, 2009; Wang, 2021). News discourse is less commonly analysed from the syntactic perspectives per se (Andriani & Bram, 2021) despite their important roles in systematizing communication messages from writers to readers. Outside the media discourse, syntactic features such as sentence types are also emphasized as a core writing skill to master in educational and professional settings, helping writers achieve communication purposes with appropriate readability and style demonstrated in their written outputs.

DuBay (2004) considers sentence length as a prominent criterion to determine readability of an English text, observing that sentence length averages have decreased over time, from 45 words per sentence during the Elizabethan times, to 29 words during the Victorian times, and about 20 words per sentence in contemporary time.

Similarly, the component of news reportage had an average sentence length of 20 words as specified in the Longman Grammar of Spoken and Written English (Biber et al., 1999b). Many researchers and experts agreed that a sentence ranging from 15 to 25 words provides comprehensible information for public communication (Russell, 2001). Moreover, Smith (2007) recommended that journalists follow an average sentence length of 20 words. Nevertheless, to many authors, syntactic elements also blend with lexis in different ways, which could complicate the ways sentences are structured. Factors such as the occurrences

and nature of clauses and subordination in a sentence and text and length of linguistic structures are potential causes of text complexity (McNamara, Crossley, & McMarthy, 2010). If the syntax of a sentence is complex, higher demands are imposed on working memory processing which causes readability challenges (ibid.).

According to Bamber and Marchman (1991), language-oriented learners become effective writers capable of forming complex sentences to show temporal and causal relationships between elements of a text and for extended development of a story theme. In educational settings, school-aged learners develop an ability to join sentences by using various linking words (Bloom, 1991). Complex sentence formation continues into the teenage years and plays a vital role in the writing skills of an individual, oral and written narrative creation, and expository text, especially in the academic setting (Reed, 1994). Furthermore, they become literate and are required to use concise syntax in written language. These tasks are not easily achievable with a simple sentence structure alone. Complex sentences allow learners to temporally and causally order events and boost the cohesiveness of narratives and increase the complexity of the written language (Haynes & Shulman, 1998; Paul, 2001).

Craig, Connor and Washington (2003) claimed that language sampling can be used as a way of reducing the impact of assumptions derived from the stereotype about learners' language use. They further suggested that the frequency of complex syntax usage be determined when assessing learners. Furthermore, they identified the use of complex syntax as skills that were positive predictors of reading comprehension of students. In terms of academic style of writing, Monippally and Pawar (2010) stated the use of complex sentence also indicates that academic writing style usually tends to use longer and complex sentences. They stated that complex sentence is the quality of sentence logic in academic prose such as journal articles. In addition, Buckingham (2015) described that journal articles have more terminologies and long noun phrases, and contain quite complex sentence structures, making language use in journal articles denser than texts for general communication purposes such as news articles.

Radford (2004) also mentioned that syntax is the component of a grammar which determines how words are combined together to form phrases and sentences. Syntax is the organization and connection between words, phrases, and clauses depending on grammatical rules that form sentences or larger constructions. Plus, Borsley (2003) stated that the syntactic analysis is key to developing precise descriptions of aspects of the syntax of various languages and the ways in which specific languages combine words to form sentences, one of the core components in human communication.

Many authors have attempted to define sentences. Hogue (1995) mentioned that a sentence is a group of words that contain a subject and a verb and expresses a complete thought. Law (2002) the most common concept is given the definition on a sentence that begins with a capital letter and finishes with a full stop, question mark or exclamation mark. One part of group word of syntax structure mentions of sentence which divided into four types of sentences such as a simple sentence, compound sentence, complex sentence, and compound-complex sentence. According to Biber et al (1999), a simple sentence is a unit that has only a single subject and a single predicate while a compound sentences two independent clauses are joined together by a coordinator or fanboys. Oshima and Hogue (2007) stated that a complex sentence contains one independent clause and one or more dependent clauses including the use of adverb clauses, adjective clauses, and noun clauses exist as part of writing complex sentences in English whist a compound-complex sentence has at least three clauses, at least two of which are independent clauses.

In terms of sentence types, Andersen (2014) that the simple sentence consists of a single subject and a single predicate forming one independent clause. The compound

sentence comprises two independent clauses. These are joined together in one sentence through a coordinator (Andersen, 2014). Radford (2009) noted that the complex sentence is made up of an independent clause which is joined by a single or multiple dependent clauses. Andersen (2014) and Raford asserted that the compound complex sentence includes aspects from both the complex and compound sentences already described above. It comprises two main clauses and at least a dependent clause.

Few studies investigating frequencies of sentence types in news discourse include those by Kongsakorn and Kongkaew (2017) on complex sentences in Thai political news written in English, Nina and Endu (2019) on Thai news articles also written in English, and Andriani and Bram (2021) on BBC news articles. No study has been conducted about sentence lengths and types of news discourse written in Khmer.

The study follows the framework of classifying Khmer sentence structures into three types: Simple, Complex, and Compound complex (ក្រុមប្រឹក្សាជាតិភាសាខ្មែរ, 2018; ធន់ ហ៊ិន, 2019; វិទ្យាស្ថានភាសាវិទ្យា, 2006) and that of Crystal and Davy (1969) and DuBay (2004) that readability of a text can be determined by sentence lengths and types.

3. Research methodology
3.1. Population and sample of the study

The articles analyzed in this study were collected from four sources, i.e., popular online newspapers in Cambodia (as indicated by numbers of daily site visitors and shares on social media such as Facebook). The articles were issued from 23 to 29 January 2020 on the topic of COVID-19 outbreak, and were archived for the study one week after they were published on the four newspapers' official websites. In total, 277 articles were collected. Since the total number of articles was large and widely dispersed, it was important to ensure that articles included in the analysis fairly represent each newspaper. Therefore, 15 articles were selected for each newspaper. To do that, two steps were taken. The first step was to identify all the tittles issued on each day. The second step was to select two titles for each day – three for the final day as it had the largest number of COVID-19-related articles. For days on which no article on COVID-19 was issued, the researchers took additional articles published on the other days to include in the target sample. Table 3.1 displays the total number of articles collected by date and by newspaper, while Table 3.2 shows the number of articles analyzed in the study by date and newspaper.

Table 3.1: Total number of articles collected listed by newspaper and date

Day issued on	Fresh News	Rasmei Kampuchea	Thmey Thmey	SABAY	Total	Selected Sample
23 January 2020	5	6	2	2	**15**	**8**
24 January 2020	7	7	1[1]	0[2]	**15**	**5**
25 January 2020	8	7	4	0	**19**	**6**
26 January 2020	13	6	5	0	**24**	**6**
27 January 2020	20	17	6	3	**46**	**9**
28 January 2020	31	17	9	18	**75**	**13**
29 January 2020	40	23	7	11	**81**	**13**
Grand Total	**124**	**83**	**36**	**34**	**277**	**60**

---

[1] For the dates no articles were published or fewer than two articles were published on Covid-19, articles were taken from the other dates as substitutes.

[2] Publication of Sabay was halted in observation of the Chinese New Year from January 24-26, 2020.

Table 3.2: Total number of articles analyzed in the study, listed by newspaper and date

Day issued on	Fresh News	Rasmei Kampuchea	Thmey Thmey	SABAY	Total
23 January 2020	2	2	2	2	**8**
24 January 2020	2	2	1	0	**5**
25 January 2020	2	2	2	0	**6**
26 January 2020	2	2	2	0	**6**
27 January 2020	2	2	2	3	**9**
28 January 2020	2	2	3	5	**12**
29 January 2020	3	3	3	5	**14**
Grand Total	15	15	15	15	**60**

3.2. Research design and instrument

To answer the research questions, both qualitative and quantitative methods were adopted. For the quantitative method, the researcher employed PAN Khmer Line Breaking Program (PKLBP) in order to separate Khmer words for word count the way English words are counted as Khmer writing conventions do not require a space in between words but phrases. The instrument would be used to obtain the primary data retrieved from four Online media in Cambodia for this study, which were then analyzed in the form of content and statistical analysis using PAN Khmer Line Breaking Program by using the standard of Dictionary of Khmer Language.

3.3. Data Collection Procedure

The researchers checked the official websites of Fresh News, Rasmei Kampuchea, Thmey Thmey, and Sabay to retrieve articles relevant to Coronavirus or COVID-19 issued in the timeframe identified above. The researchers looked up articles by using such key words as កូរ៉ូណា កូវ៉ីណា កូរូណា on the entire websites. All articles were copied and pasted on the Microsoft word program for archiving as a text database for analysis later.

3.4. Data Analysis and Statistical Procedure

After collecting all articles related to Coronavirus, the researchers used the PAN Line Breaking Program to separate words by zero space. The researchers then used the "find and replace" in Microsoft word program by using the Paragraph Mark to count the words and sentences. No text in the pictures, headlines, captions, name of authors and date of publication were counted and analyzed. The data for this study was analyzed using the SPSS version 25 computer software program. Descriptive statistics was utilized to analyze the data for this study, providing information on frequency distribution, mean, standard deviation, and range to count words, sentences and paragraphs. Plus, word count was also conducted at the sentence, paragraph, and text levels, and sentence count at the paragraph and text levels. Number of paragraphs per text was also counted. Analysis of sentence types was based on classification of Khmer sentences in. To ensure a high level of reliability in classifying Khmer sentences into the three types, the researchers each analyzed the sentences separately, and then compared their responses. Where differences occurred, an extensive discussion was conducted to reach an agreement, on several occasions with an intervention from a fourth analyst, who is an established Khmer linguist on Zoom web conferences. Counting sentences were defined by a full spot in Khmer (។), exclamation mark (!) and question mark (?). For quoted sentences, their numbers were tallied, but their sentence types were not analyzed as quoted speeches are not produced by the writers themselves and may contain multiple sentence types within one quotation.

In terms of interpretation, the scale of readability formulae are mathematical equations whose results provide interpretations for the description of a text from 'very easy'

to 'very difficult'. They are meant to help predict or indicate reading ability in terms of school grade level required to read and comprehend a particular written text (DuBay, 2004). The formulations are based on research results that classify vocabulary challenges measured in word comprehension or sophistication, or word length and sentence structure determined as two accurate variables used in this study of text complexity and difficulty in sentence length (DuBay, 2004).These features are equally prominent in measuring or assessing text complexity in applied linguistics and legal studies (Lu, 2010; McNamara, Crossley, & McMarthy, 2010). The most frequently cited and used formulae to date are Flesch Reading Ease (1948), DuBay (2004) and Wikidot (2020). The average sentence length, estimated school grade completed and readability were formulated in Table 3.3.

Table 3.3: Average sentence length, estimated school grade completed, and readability Flesch Reading Ease (1948), DuBay (2004) and Wikidot (2020)

Average sentence length in words	Estimated school grade completed	Readability
8 or less	4th grade	Very easy
9	5th grade	Easy
14	6th grade	Fairly easy
17	7th or 8th grade	Standard
21	Some high school	Fairly difficult
25	High school or some colleges	difficult
29 or more	college	Very difficult

4. Research findings

Results from this study are divided into two sections. The first section analyzed descriptive statistics as mean, standard deviation, maximum, and minimum and inferential statistic through tests such as independent T-test, One-way ANOVA, and Pearson correlation coefficient on the average length of an article, paragraph, and sentence in the four popular online newspapers in Khmer, in comparison to academic texts. These results will answer the first research question. For the second section, descriptive statistics were used to answer the second research question by analyzing the most commonly used types of sentences among the articles collected from the four newspapers.

Before presenting the analyses for the two research questions, here demographic data related to the four popular online newspapers are presented, related to the number of articles, paragraphs, sentences, and overall words. A total of four online newspapers were conducted in this study. Table 4.1 presented a summary of the demographic figures of the newspapers collected in the valid action of retrieving online newspapers.

Table 4.1: Total number of titles, paragraphs, sentences, and repeated words analyzed in the study

Newspapers	Articles	Paragraphs	Sentences	Non-repeated words	Overall words
Thmey Thmey	15	69	130	646	3,343
Sabay	15	48	93	752	2,742
Fresh News	15	66	101	925	4,337
Rasmei Kampuchea	15	91	118	1,148	3,896
Grand Total	**60**	**274**	**442**	**2,488**	**14,318**

The demographic variables of online newspapers in terms of the number of titles, paragraphs, sentences and repeated words. The number of titles was equivalent to Thmey Thmey, Sabay, Fresh News and Rasmei Kampuchea online newspapers as 15 while the

number of grand total paragraphs was 274. Thmey Thmey and Fresh News were similar among 69 and 66 paragraphs and the maximum of paragraphs were 91 of Rasmei Kampuchea whilst the minimum of paragraphs was 48 of Sabay. The grand total of the sentences of four online newspapers were 442 covering 93 of Sabay, 101 of Fresh News, 118 of Rasmei Kampuchea and 130 of Thmey Thmey, individually. The grand total of non-repeated words was 2,488 words among Thmey Thmey, Sabay, Fresh News, and Rasmei Kampuchea. The most non-repeated words were Rasmei Kampuchea of 1,148 words, whilst the lowest non-repeated words were Thmey Thmey of 646 words. The Fresh News and Sabay were 925 and 752 non-repeated words, respectively. The grand total of overall words was 14,318, which Rasmei Kampuchea and Fresh News were similarity proportion of 3,896 and 4,337, and which Sabay and Thmey Thmey were 2,742 and 3,343 overall words.

4.1. Results for the number of titles, paragraphs, sentences, and repeated words

The first question aims to investigate the average length of an article, paragraph, and sentence among four popular online newspapers in Khmer, in comparison to academic texts. The scale of readability formulae is mathematical equations whose results provide interpretations for the description of a text from 'very easy' to 'very difficult'.

As shown in the Table 4.2, four online newspapers as Thmey Thmey, Sabay, Fresh News and Rasmei Kampuchea including number of words in an article, number of paragraphs, number of sentences, number of words of a paragraph, and number of words in a sentence in an article were analyzed.

Table 4.2: Four newspapers analyzed in the study

Four newspapers	N	Min	Max	Mean	S.D
Thmey Thmey					
Number of words in an article	15	69	532	222.87	141.022
Number of paragraphs in an article	15	2	10	4.60	2.44
Number of sentences in an article	15	3	20	8.67	5.04
Number of words in a paragraph	15	23	67.5	47.21	11.71
Number of words in a sentence	15	19.17	33.8	25.34	3.51
Sabay					
Number of words in an article	15	88	255	182.80	43.80
Number of paragraphs in an article	15	2	4	3.20	0.56
Number of sentences in an article	15	3	8	6.20	1.70
Number of words in a paragraph	15	34	85	58.11	16.61
Number of words in a sentence	15	12.57	61.33	31.89	12.09
Fresh News					
Number of words in an article	15	96	824	289.13	178.20
Number of paragraphs in an article	15	3	9	4.40	1.55
Number of sentences in an article	15	3	16	6.73	3.45
Number of words in a paragraph	15	32	91.56	63.22	20.96
Number of words in a sentence	15	28	63.8	42.11	10.05
Rasmei Kampuchea					
Number of words in an article	15	107	566	259.73	115.73
Number of paragraphs in an article	15	3	10	6.07	2.28
Number of sentences in an article	15	3	15	7.87	3.20
Number of words in a paragraph	15	27.33	81.25	43.75	16.30
Number of words in a sentence	15	25.09	51.25	33.66	7.83
Grand total					
Number of words in an article	**60**	**69**	**824**	**238.63**	**132.30**

Four newspapers	N	Min	Max	Mean	S.D
Number of paragraphs in an article	60	2	10	4.57	2.09
Number of sentences in an article	60	3	20	7.37	3.59
Number of words in a paragraph	60	23	91.56	53.07	18.13
Number of words in a sentence	60	12.57	63.8	33.25	10.61

The smallest total number of words in an article is 69 words, while the biggest is 824 words, with a mean value of M = 238.63 and standard deviation (SD) of 132.30. For paragraphs per article, the smallest number is 2, while the highest is 10, with a mean value of M = 4.57 and standard deviation (SD) of 2.09. For numbers of sentences per article, the smallest number is 3 while the highest is 20 (M = 7.37 and SD = 3.59).

For word counts, words per paragraph average at 53.07 (M = 53.07, S.D = 18.13), while an average sentence length was found to be at 33.25 (SD = 10.61).

The table 4.3 depicts the summary of sentence word counts for Thmey Thmey, Sabay, Fresh News and Rasmei Kampuchea.

Table 4.3: Total number of words in a sentence analyzed in the study

Number of words in a sentence	Min	Max	Mean	Std. Deviation
Thmey Thmey	19.17	33.8	25.34	3.51
Sabay	12.57	61.33	31.89	12.09
Fresh News	28	63.8	42.11	10.05
Rasmei Kampuchea	25.09	51.25	33.66	7.83
Grand Total	12.57	63.8	33.25	10.61

These results suggest that the readability and comprehensibility of the sampled articles seem rather low and the four popular online newspapers appeared to appeal to college students if Table 4.3 is to go by. The average length of Fresh News sentences was the highest (M = 42.11, SD = 10.05), while Thmey Thmey had the shortest sentences (M = 25.34, S.D = 3.51). The other two newspapers had similar average sentence lengths – with M = 31.89, SD = 12.09 for Sabay and M = 33.66, S.D = 7.83 for Rasmei Kampuchea.

Figure 4.1 presents an average number of words per sentence in Khmer guide books, Thmey Thmey, critical thinking textbook, Khmer grade 12 textbook, History textbook for higher education, Sabay, Rasmei Kampuchea, and Fresh News. The highest mean of words in a sentence was found for Khmer grade 12 (44.58 words) whilst the lowest mean was found for the Sangapac Critical thinking (27.22 words). The average numbers of words in a sentence for Thmey Thmey, Resmei Kampuchea, Sabay and guide book were similar, averaging between 31.89 of 36.19 words. History textbook for higher education and Fresh news sentence lengths hover at around 36 words per sentence.
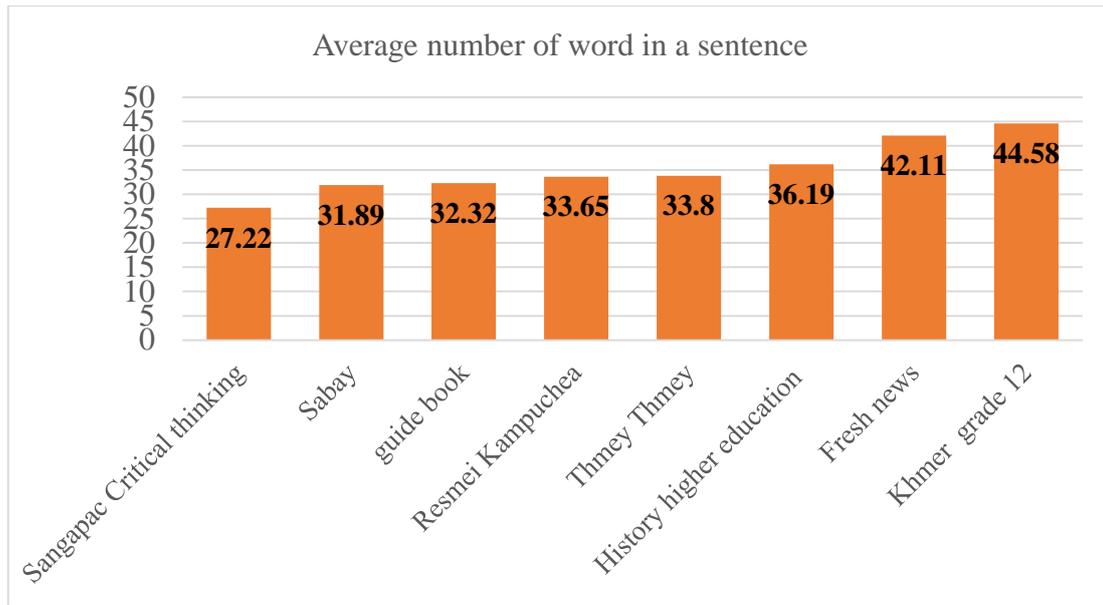
Figure 4.1: Average number of words in a sentence

The researchers conducted an analysis of variance (ANOVA) of the number of words, paragraphs, and sentences per article, and the number of words in a sentence for the four online newspapers. The numbers of words and sentences in an article had no statistically significant difference among all the texts included in the analysis at the 95% confidence level (F [3, 59] = 1.9, $p = 0.14$ for sentences across articles and F [3, 59] = 1.467, $p = 0.233$ for words across articles). However, the number of paragraphs per article and words in a sentence had a statistically significant difference at the 0.5 (F [3, 59] = 5.968, $p = 0.01$) and (F [3, 59] = 8.933, $p = 0.00$). This result showed that the numbers of paragraphs and the numbers of words in a sentence written in the four online newspapers were of different quantity, as clarified in Table 4.4.

Table 4.1: Differences and Similarities in lengths among the articles from the four newspapers

Four types of Newspapers					
	Df	SS	MS	F	Sig.
Number of words in an article					
Between Groups	3	95421.133	31807.044	1.9	0.140
Within Groups	56	937318.8	16737.836		
Total	59	1032739.933			
Number of paragraphs in an article					
Between Groups	3	62.2	20.733	5.968	0.010
Within Groups	56	194.533	3.474		
Total	59	256.733			
Number of sentences in an article					
Between Groups	3	55.533	18.511	1.467	0.233
Within Groups	56	706.4	12.614		
Total	59	761.933			
Number of words in a sentence					
Between Groups	3	2147.806	715.935	8.933	0.000
Within Groups	56	4488.238	80.147		
Total	59	6636.044			

p< 0.05

Table 4.5 displays the results of correlation among of number of sentences per article, number of words per article, number of paragraphs per article, and number of words per sentence of the four online newspapers. The results presented that there is statistically significant positive correlation coefficient between them, ranging from direct and moderate (Pearson $r = 0.352$, $p < 0.01$) to direct and very strong (Pearson $r = 0.826$, $p < 0.01$). Therefore, the null hypothesis was rejected. However, the analyses suggested that the correlation coefficient was not significant between the number of words per sentence and the number of sentences per article (Pearson $r = -0.168$, $p > 0.05$), nor between the number of words per sentence and the number of paragraphs per article (Pearson $r = 0.049$, $p > 0.05$).

Table 4.5: Correlation among number words in number of sentences, number of words in an article, number of paragraphs, and grand total words in sentences

		Number of sentences per article	Number of words per article	Number of paragraphs per article	Number of words per sentence
Number of sentences per article	Pearson Correlation	1			
	Sig. (2-tailed)				
	N	60			
Number of words per article	Pearson Correlation	.741**	1		
	Sig. (2-tailed)	0			
	N	60	60		
Number of paragraphs per article	Pearson Correlation	.826**	.770**	1	
	Sig. (2-tailed)	0	0		
	N	60	60	60	
Number of words per sentence	Pearson Correlation	.352**	0.049	-0.168	1
	Sig. (2-tailed)	0.006	0.712	0.198	
	N	60	60	60	60

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4.6 displays the different means of sentence lengths found in the target newspaper articles and academic texts. The average sentence length of the newspaper articles (M=33.25, SD = 10.60) was not significantly shorter than the that of academic texts (M = 34.30, SD = 21.02) at the .05 level ($t = -0.35$, df = 123). According to this result, there was no significant difference in the average number of words per sentence in newspaper articles and academic texts (F = 2.049, $p = 0.155$). Since no statistically significant mean difference found in the sentence lengths of newspaper articles and academic texts, the null hypothesis was maintained, suggesting that these news articles had the same level of readability as that of academic texts for senior high school and university students.

Table 4.6: Difference between newspapers and academic texts

Differences between newspapers and academic texts							
	Newspapers (N = 60)		Academic texts (N = 67)		F	t	p-value
	M	SD	M	SD			
Number of words per sentence	33.25	10.60	34.30	21.02	2.049	-0.35	0.155

$p < 0.05$

4.2. Results for the most commonly used types of sentences

The second research question examined in this study is: "What are the most commonly used types of sentences among the articles collected from the four newspapers?" This question is to investigate the most commonly used types of sentence structure in the popular online newspapers since this may provide an insight into the ways these journalistic authors chose to convey their messages in the online newspaper articles. Descriptive statistics was used to analyze the mean, standard deviation, maximum, and minimum in the sampled articles included in the study. Before using descriptive statistics, as explained in the previous section, the researchers analyzed the 444 sentences separately, and then compared their responses. Where differences occurred, an extensive discussion was conducted to reach an agreement, on several occasions with an intervention from a fourth analyst, who is an established Khmer linguist via Zoom web conferences.

As presented in Table 4.7, the number of quotations, simple sentences, complex sentences, and compound-complex sentences were analyzed in the 60 articles retrieved from four online newspapers.

Table 4.7: Four newspapers analyzed the quotation, simple, complex, and complex-compound sentences

Four newspapers	N	Min	Max	Mean	S.D
Thmey Thmey					
Number of quotations in an article	6	1	3	1.50	0.84
Number of simple sentences in an article	13	1	6	3.31	1.89
Number of complex sentences in an article	14	1	9	4.14	2.57
Number of compound-complex sentence in an article	12	1	4	1.67	0.99
Sabay					
Number of quotations in an article	1	1	1	1.00	.
Number of simple sentences in an article	13	1	4	2.15	1.14
Number of complex sentences in an article	15	1	6	3.40	1.24
Number of compound-complex sentence in an article	8	1	3	1.63	0.74
Fresh News					
Number of quotations in an article	3	1	4	2.00	1.73
Number of simple sentences in an article	10	1	3	1.70	0.82
Number of complex sentences in an article	12	1	7	2.50	1.88
Number of compound-complex sentence in an article	13	1	8	3.69	2.14
Rasmei Kampuchea					
Number of quotations in an article	2	1	2	1.50	0.71
Number of simple sentences in an article	12	1	6	2.25	1.55
Number of complex sentences in an article	15	1	10	3.67	2.29
Number of compound-complex sentence in an article	15	1	4	2.20	0.94
Grand total					
Number of quotations in an article	**12**	**1**	**4**	**1.58**	**1.00**
Number of simple sentences in an article	**48**	**1**	**6**	**2.40**	**1.51**
Number of complex sentences in an article	**56**	**1**	**10**	**3.46**	**2.08**
Number of compound-complex sentence in an article	**48**	**1**	**8**	**2.37**	**1.57**

The most commonly used were complex sentences of 56 ranging the minimum number of 1 to the maximum number of 10 sentences in an article and mean (M = 3.46, S.D

= 2.08) while the lowest figures of the types of sentences were the number of quotation sentence of 12 consisting of mean (M = 1.5, S.D = 1.00) and starting the minimum proportion of 1 to the maximum proportion of 4. The equivalent number of simple and compound-complex sentences was 48 articles including the mean of simple sentence (M = 2.40, S.D = 1.51) and the mean of compound-complex sentence (M = 2.37, S.D = 1.57) and the same lowest volume of 1 to the highest volume of 6 of simple sentence and 8 compound-complex sentence. Furthermore, the types of sentences were analyzed in 60 articles following Khmer syntax such as quotation sentences, simple sentences, complex sentences, and compound-complex sentences. The maximum number of quotation sentences, simple sentences, complex sentences, compound-complex sentences were 6 articles of Thmey Themy, 13 articles of Thmey Thmey and Sabay, 15 articles of Sabay and Rasmei Kampuchea, and 15 articles of Rasmei Kampuchea, respectively. In contrast, the lowest number of quotation sentences, simple sentences, complex sentences, and compound-complex sentences were 1 article of Sabay, 10 and 12 articles of Fresh News, and 8 articles of Sabay, separately.

As shown in Figure 4.2, types of sentences in the 60 articles of the target online newspapers were analyzed. The least common type was the quotation (19, which is equal to 4%) while the most common one was complex sentence (194, which is equal to 44%). 114 and 115 simple and compound-complex sentences were used, respectively, which are equal to 26%, approximately. The fact that a large proportion of complex sentences was used by journalists in the four online newspapers could be attributed to a number of possible reasons discussed in the next section.
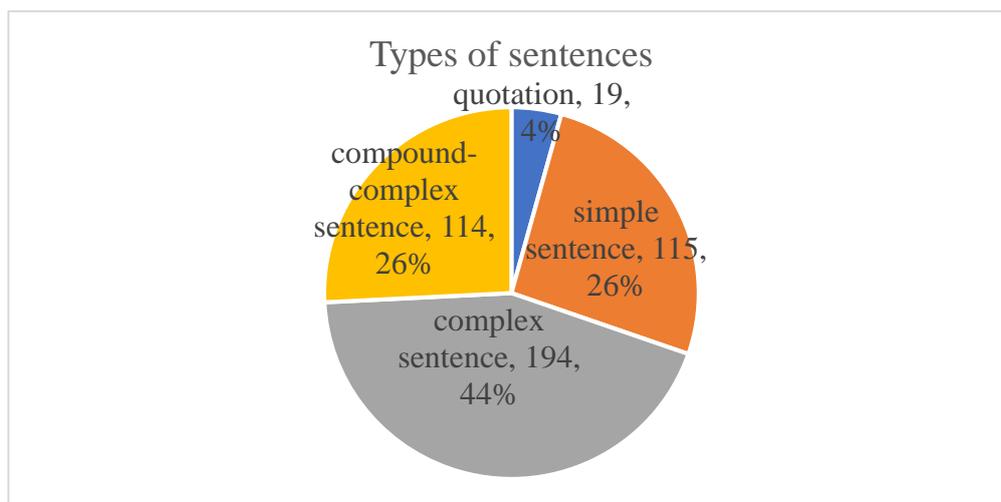

Figure 4.2: Average number of words in a sentence

4. Discussion and conclusion

This research study sets out to conduct a syntactical analysis, in particular sentence structure, of 60 articles published in four online newspapers written in Khmer in Cambodia. It tries to answer two main research questions specified in Section 1. As explained in the findings section, the average number of words in a sentence in academic text and four popular online newspaper was not significantly different at the .05 level (t = -0.35, df = 123). The finding showed that the newspapers and academic texts had similar readability and comprehensibility levels and hereby appear to be written for college students. These long sentences are likely due to the trend among Cambodian journalistic writers to use multiple nominal groups and clauses within each sentence, either because of their under-using sentence-breaking punctuation marks and insufficient trainings in written media discourse

conventions. The results from this study are consistent with the statements by McNamara et al. (2010) who mentioned that potential causes of text complexity include factors such as the frequency and function of clauses and subordination in the sentence and text and length of linguistic structures. They also explained that readability challenges can occur if the syntax of a sentence is complex, and higher demands are placed on working memory processing. The study findings contrast suggestions by DuBay (2004) and Biber, Johansson, Leech, Conrad, and Finegan (1999b) who claimed that news reportage had an average sentence length of 20 words as cited in Longman Grammar of Spoken and Written English. Russell (2001), in fact, suggested that a sentence containing between 15 to 25 words is most suitable and comprehensible for public communication, while Smith (2007) recommended that journalists refrain from going beyond an average of 20 words per sentence.

As mentioned in second research finding, complex sentence was most commonly used in the target online newspapers in line with the findings by Kongsakorn and Kongkaew (2017), Nina and Endu (2019), and Andriani and Bram (2021). The results showed that complex sentence is easy to be extended with added syntactic categories such as clauses and phrases. The findings suggest that the writers for the four newspapers had an ability to connect sentences, forming complex relationships within and between clauses the way academic text is written. The results from this study are consistent with the statements by Bamberg and Marchman (1991), Haynes and Shulman (1998) and Paul (2001) who stated that complex sentences allow language users to temporarily and causally order events and boost the cohesiveness of narratives and increase the complexity of the written language. Plus, Monippally and Pawar (2010) supported the use of complex sentence especially for academic writing style which tends to favour longer and complex sentences. However, Reed (1994) and Buckingham (2015) argued that if the journal articles have more terminologies, long noun phrases, and quite complex structures, the language used in journal articles becomes denser.

To conclude, the study shows that Khmer journalistic sentence structures follow those of English news discourse. However, the average sentence length appeared higher than those recommended by international literature which could potentially hinder readability and comprehensibility from the readers' perspectives. One main reason longer sentences in Khmer newspaper writing could be adherent in the contemporary Khmer writing style which uses fewer punctuation marks, as compared to prescriptive grammar rules specified in the academic Khmer grammar textbooks. The other important implication pertains to the role of editors in ensuring the written outputs of journalist writers are appropriate for their target audience. During the specified publication period, information about COVID-19 outbreaks, preventive measures, and treatment was relatively scarce, and there was a high demand of such information, possibly prompting the editors to approve articles for publication in the four newspapers with less editing of language use and sentence structures than normal.

On the pedagogical and professional training fronts, these four newspapers are popular on social media, of which a large proportion of the audience is teenagers, rather than higher education students. Continuous professional training in stylistic writing without overdoing of sentence combination and/or expansion while focusing on readability and comprehensibility will be indispensable in helping writers and editors close this gap.

For future studies, it is recommended that the such an analysis be expanded to include other grammatical features at the sentence level to develop a better understanding of how these features interact with sentence lengths and structures, and possibly to be connected to semantic and pragmatic analyses so as to factor in adherent meanings of sentences and those intended by the writers themselves. Such studies will benefit from an attempt to seek direct documentation of the writers' intended meaning, for example, via an interview or diary entries, as opposed to those perceived by the target readers.

References

Andersen, S. (2014). *Sentence types and functions*. San Jose State University Writing Center.

Andriani, E., & Bram, B. (2021). Syntactic analysis of sentence patterns and types of BBC news articles. *LiNGUA: Jurnal Ilmu Bahasa Dan Sastra*, *16*(1), 25–34.

Bamberg, M., & Marchman, V. (1991). Binding and unfolding: Towards the linguistic construction of narrative discourse. *Discourse Processes*, *14*(1), 277–305.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999a). *Longman grammar of spoken and written English*. Pearson Education Limited.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999b). *Longman grammar of spoken and written English*. Pearson Education Limited.

Bittner, J. R. (2003). *Mass communication: An introduction* (5th ed.). Heinemann.

Bloom, L. (1991). *Language development from two to three*. Cambridge University Press.

Borsley, R. D. (2003). *Syntactic theory a unified approach* (2nd ed.). Arnold.

Buckingham, L. (2015). *Doing a research project in English studies: A guide for students*. Routlegde.

Chen, Y. (2017). A critical discourse analysis of news reports on Sino-Japan Boat Collision. *DEStech Transactions on Social Science, Education and Human Science*, 135–139. https://doi.org/10.12783/dtssehs/eeres2016/7633

Craig, H. K., Connor, C. M., & Washington, J. A. (2003). Early positive predictors of later reading comprehension for African American students: A preliminary investigation. *Language, Speech, and Hearing Services in Schools*, *34*(1), 31–43.

Crystal, D., & Davy, D. (1969). *Investigating English Style*. Longman.

DuBay, H. W. (2004a). *The principles of readability*. Impact Information.

Fairclough, N., & Wodak, R. (1997). Critical discourse analysis. In *In: Van Dijk (T.). Discourse as Social Interaction*. Sage.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(1), 221–233.

Grossberg, L., Wartella, E., & Whitney, D. (1998). *Media making: Mass media in a popular culture*. Sages.

Hassan, A. (2018). Language, media, and ideology: Critical discourse analysis of Pakistani news bulletin headlines and its impact on viewers. *SAGE Open*, *8*(3), 1–15. https://doi.org/10.1177/2158244018792612

Haynes, W. O., & Shulman, B. B. (1998). *Communication development: Foundations, processes, and clinical application*. Williams & Wilkins.

Hogue, A. (1995). *Academic writing*. Longman.

Järvbäck Hillbom, K. (2009). *Newspaper Readability: A Broadsheet vs. a Tabloid*.

Kongsakorn, J., & Kongkaew, P. (2017). An analysis of the complex sentences used in selected Thai political news in online Bangkok Post. *International Journal of Research-GRANTHAALAYAH*, *5*(3), 16–28.

Law, J. (2002). *The language Toolkit: Practical advice on English grammar and usage*. Oxford University Press.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474–496.

McNamara, D. S., Crossley, A. S., & McMarthy, M. P. (2010). Linguistic features of writing quality. *Written Communication*, *27*(1), 57–86.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, *27*(1), 57–86.

Monippally, M. M., & Pawar, B. S. (2010). *Academic writing: A guide for management students and researchers*.

Nina, M. G., & Endu, M. U. (2019). Analisis Kesulitan Siswa SMP dalam Menyatakan Peristiwa Sehari-Hari dan Menurunkan Rumus Luas Bangun Datar Segiempat. *JPMI (Jurnal Pembelajaran Matematika Inovatif)*, *2*(1), 51–56.

Nkemdilim, P. (2015). Assessment of Patterns of Readership of Online Newspapers in Selected Nigerian Universities. *International Journal of Linguistics and Communication*, *3*(2), 35–46.

Okonofua, G. (2010). Prospect and issues of online newspapers in Nigeria. *A Paper Presented at the African Council for Communication Education Conference Held in American University of Nigeria, Yola, Adamawa State.*

Oshima, A., & Hogue, A. (2007). *Introduction to academic writing.* Pearson Education.

Pang, H., & Wu, S. (2009). Critical instance analysis of news English discourse. *English Language Teaching, 2*(2), 148–151. https://doi.org/10.5539/elt.v2n2p148

Paul, R. (2001). *Language disorders from infancy through adolescence* (2nd ed.). Mosby.

Radford, A. (2004). *An introduction to English sentence structure.* Cambridge.

Radford, A. (2009). *An introduction to English sentence structure.* Cambridge university press.

Reed, V. (1994). *An introduction to children with language disorders* (2nd ed.). Merrill.

Russell, S. (2001). *Grammar, structure & style: A practical guide to advanced level English language.* Oxford University Press.

Smith, J. (2007). *Essential reporting: The NCTJ guide for trainee journalists.* Sage.

Wang, Y. (2021). A critical discourse analysis of news reports on Covid-19 in people's daily and the New York Times. *Jurnal IPTEK-KOM (Jurnal Ilmu Pengetahuan Dan Teknologi Komunikasi), 23*(1), 49–62. http://dx.doi.org/10.33169/iptekkom.23.1.2021.49-62

Wikidot. (2020). *Readability guidelines—Alpha:Universal style guide.*

ក្រុមប្រឹក្សាជាតិភាសាខ្មែរ. (2018). *សៀវភៅវេយ្យាករណ៍ភាសាខ្មែរ.* ក្រុមប្រឹក្សាជាតិភាសាខ្មែរ.

ធន់ ហ៊ិន. (2019). *វេយ្យាករណ៍ ជួន ណាត.* គ្រឹះស្ថានឯកសារខ្មែរ.

វិទ្យាស្ថានភាសាវិទ្យា. (2006). *វេយ្យាករណ៍ទម្រង់និយមនៃភាសាខ្មែរទំនើប.* វិទ្យាស្ថានភាសាវិទ្យា.