

Comparing the Lexical Density of Spoken and Written News

Mulyati Khorina ^{a,1,*}, Wulandari Zahara Handani ^{b,2}

^{ab} English Department, Politeknik Negeri Bandung, Bandung, Indonesia

¹ Email: mulyati.khorina@polban.ac.id ; ²wulandari.zahara.bing19@polban.ac.id

*Corresponding author

ARTICLE INFO

Article history:

Received: 30/07/2022

Revised : 09/08/2022

Accepted: 16/12/2022

Keywords:

lexical density

spoken news

written news

lexical items

noun

ABSTRACT

This paper is concerned with lexical density in spoken and written news on BBC News. The purpose of this study is to investigate the lexical density found in spoken and written news regarding the Afghanistan and Taliban conflict broadcasted by BBC. The method applied in this study was a qualitative method. The data consist of five spoken news taken from the BBC YouTube channel and five written news obtained from the BBC news website. The results show that generally, the average lexical density levels of both spoken and written news tend to be the same, that is, higher than 50%. Specifically, spoken news has an average lexical density level of 52.15% while written news has an average lexical density of 55.05%. Meanwhile, the lexical items which influence the density of the texts are nouns, verbs, adjectives, and adverbs with the noun to be dominant.

I. Introduction

This paper is concerned with the lexical density of spoken and written news. The lexical density is the proportion of content words (lexical items) in a given discourse which can be measured either as the ratio of lexical items to a total number of words or as the ratio of lexical items to the number of higher structural items in the sentences (for example, clauses)[1][2]. Two formulas are available to calculate lexical density. Ure's (1971) calculation is based on the proportion of the number of content words over the total number of words times 100. Whereas Halliday [3] changed the total number of words as the denominator to the total number of clauses in the text. Besides, Ure (1971) in [4] discriminated the lexical density of spoken texts from written ones. The former generally has a lexical density under 40% while the latter has a lexical density higher than 40%. Khamahani [5] who investigated the lexical density of Azeri News and Tehran Times headlines found out that both of the news is highly dense in which Azeri News has a lexical density of 88% which is higher than the ones of Tehran Times (75%). It means that the study corresponds with Ure's study. Since there is a significant difference between the lexical density of spoken and written texts, this study analyzed the lexical density of spoken and written news broadcasted by BBC News.

Identifying the lexical density of text may have some purposes. First, it may be used to find out the complexity of a textbook used for English language teaching. By doing so, the textbook can be evaluated whether it is suitable for the level at which it is used [6]–[9]. Also, lexical density can be used as a parameter to assess students' writing assessment. Through lexical density analysis, the progress of students' language learning, especially in writing can be found since the result may show their ability in writing an academic text, such as the introduction section of thesis proposals written by English graduate students. When the lexical density of students' writing is 51.19%, it is considered less dense meaning the students still have limited knowledge about the language used in writing an academic text implying that the student's language knowledge needs to improve to make their work meet the standard of academic language [10].

Based on what has been presented above, this study focused on the analysis of the lexical density of spoken and written news broadcasted by BBC News concerning the Taliban overtaking the Afghanistan government in August 2022. Thus, the research questions formulated as below.

- What is the lexical density of spoken and written news?
- What content words found in both types of the news?

II. Literature Review

A. News

Getting information, such as news in this modern life is essential. As defined by Ekstrom [11], news is "reliable, neutral and current factual information that is important and valuable for citizens in democracy". Moreover, Fowler [12] elaborated that "News is not a natural phenomenon emerging straight from 'reality', but a product. It is produced by an industry, shaped by the relations between the media and other industries, by the bureaucratic and economic structure of that industry, and most importantly by relations with government and with other political organizations." In other words, news is information about current events in a specific area, country, or world that is published in a newspaper, broadcast on radio, shown on television, and reported on the Internet. It plays an important role in modern society and serves as a window through which people gain knowledge about the world due to its benefits such as timeliness and objectivity. The information provided by news reports serves to guide the public [13]. Accordingly, getting reliable information about what is happening in a particular region from any reliable sources, written or spoken, is critical for citizens in democracy. For example, the news about the Taliban in Afghanistan in 2021 which was a trending topic in the world aired by all broadcasting companies in the world.

B. Ure's lexical density

Lexical density refers to a concept in computational linguistics that measures the structure and complexity of human communication in a language [1]. The estimation of linguistic complexity in a written or spoken texts can be seen from the functional words (grammatical units) and content words (lexical units, lexemes). One way to calculate the lexical density is to compute the ratio of lexical items to the total number of words as proposed by Ure (1971) in [4]. While Halliday [3] used total number of clauses as the denominator. The formula used by Ure is as follows.

$$Ld = \frac{\text{The number of lexical items} * 100}{\text{The total number of words}}$$

For example: *Their **lightning advance prompted tens of thousands of people to flee their homes, many arriving in the Afghan capital, others heading for neighbouring countries.*** (<https://www.bbc.com/news/world-58232525>)

Applying Ure's formula above it can be obtained that the lexical density (LD) of the text above is:

$$LD = \frac{16 * 100}{24}$$

$$LD = 66,67$$

In the text above, there are 16 lexical items or content words while the total number of words in the text is 24. Therefore, the total number of lexical items (in bold) is divided by the total number of words times 100 to get the LD which is 66,67.

Linguistically, the lexical items or content words refer to words that possess semantic content and contribute to the meaning of the sentence in which they occur. According to Radford [14], lexical items are also called lexical categories which include noun, verb, adjective, adverb, and preposition. However, in this paper preposition is excluded from the list since it does not have semantic content when it stands alone. Thus, in this paper, the lexical items which become the nominator of the formula are noun, verb, adjective, and adverb.

I. Method

This study applied a qualitative method. The data were obtained from bbc.com for written texts and YouTube of BBC for spoken one as listed below. The data collected were only those which were published and broadcasted in August 2021. The data used in this study consist of five texts of spoken news taken from YouTube of BBC. These data were transcribed first before analyzing. Whereas, the data of written news consist of five texts too. All the data were listed in Table 1. And the steps of analysis were illustrated in Figure 1.

Table 1. List of the data collected

No	News Title	Type of News	Word Count	Link
1	How the Taliban Stormed across Afghanistan in 10 Days	Spoken	104	https://www.youtube.com/watch?v=xJivUISR6As
2	How the Taliban Stormed across Afghanistan in Ten Days	Written	604	https://www.bbc.com/news/world-58232525
3	Life under Taliban Rule One Month on	Spoken	343	https://www.youtube.com/watch?v=nbxZjNWEqaU&t=203s
4	Afghanistan: Life under Taliban Rule One Month on	Written	850	https://www.bbc.com/news/world-asia-58550640
5	US and Taliban Hold First Talks since Afghanistan Withdrawal	Spoken	557	https://www.youtube.com/watch?v=TJqMJENtBH8
6	Afghanistan: US and Taliban Hold First Face-to-Face Talks since Withdrawal	Written	391	https://www.bbc.com/news/world-asia-58854984
7	Afghanistan's Booming Drugs Trade under Taliban Rule	Spoken	499	https://www.youtube.com/watch?v=QcX8a0MSdhY&t=96s
8	Meth and Heroin Fuel Afghanistan Drugs Boom	Written	1315	https://www.bbc.com/news/world-asia-59608474
9	'No Work and No Money': Afghans Settle into Life under Taliban Rule	Spoken	373	https://www.youtube.com/watch?v=6LqxBAo1lFw
10	'There's Security but No Money': Afghans Settle into Life under Taliban Rule	Written	855	https://www.bbc.com/news/world-asia-58597432

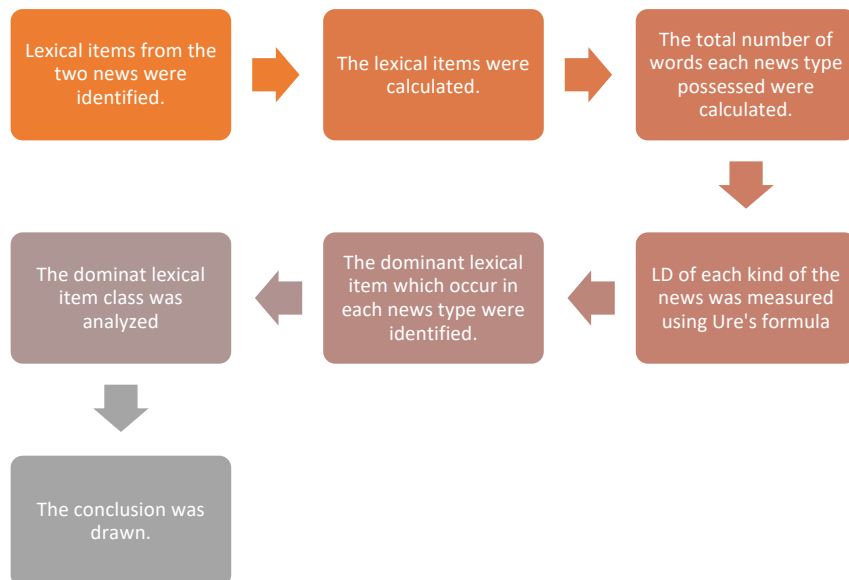


Figure 1. The steps of analysis

III. Results and Discussion

A. Results

Figure 2. shows that in general both spoken and written news have a lexical density higher than 50%. The findings only confirm Ure’s for the lexical density of written news. On the other hand, the finding of the lexical density owned by spoken news was not approved.

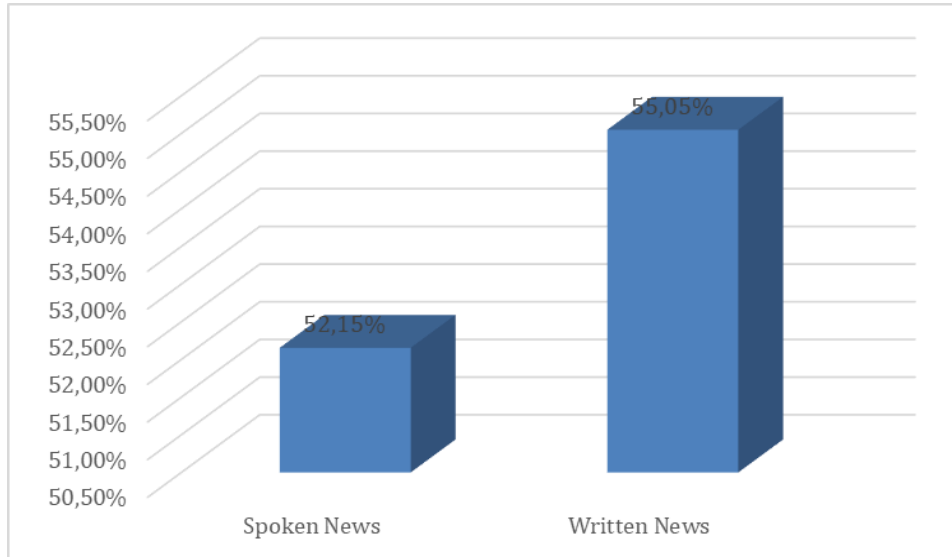


Figure 2. The Lexical Density Level of Spoken and Written News

In detail, the data of lexical density levels of both spoken and written news can be found in the table below. Most of spoken news and written news have lexical density levels higher than 50% except one of spoken and written news whose lexical density levels are below 50%. The highest lexical density level found is 59,36% which much lower than Khamahani’s [5] findings of 75% and 88%. However, the lexical density level of spoken news which is generally above 50% does not correspond with Ure’s which stated that spoken texts has lexical density below 40%.

Table 2. The individual lexical density levels of the spoken and written news

News No.	Spoken News	Written News
1	53.06%	55.06%
2	50.96%	57.95%
3	46.68%	59.36%
4	54.31%	54.22%
5	55.76%	48.65%
Average	52.15%	55.05%

Furthermore, the lexical items which influence the density of the texts are noun, verb, adjective, and adverb. In spoken news and written news, nouns to dominate the lexical items which make the texts dense. But the number of nouns found in written news seem to be double compared with the one in spoken news. Another lexical item, verbs, occur almost half of nouns, but still they are higher in written news than in spoken one. The other lexical items, adjectives and adverbs also occur higher in written news than in spoken ones. Yet, the number of their occurrences are approximately twice as many as in spoken news as seen in Figure 3.

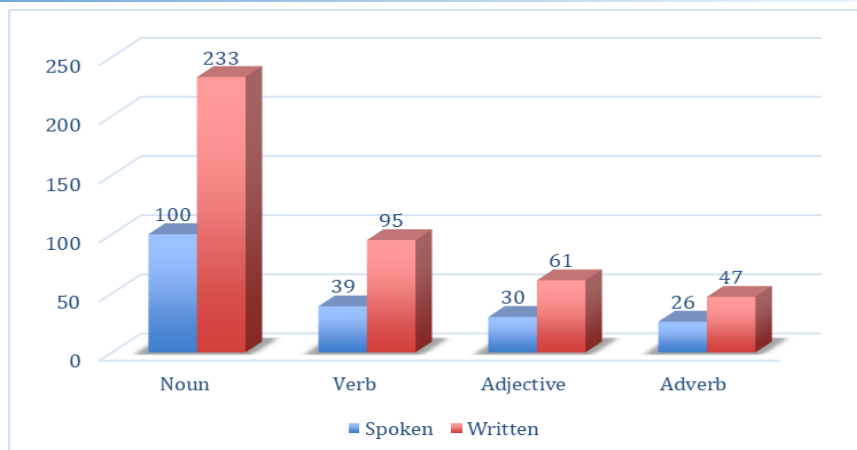


Figure 3. The lexical items which occur in spoken and written news

B. Discussion

Referring to Figure 3. The dominant lexical item found in the data is noun either in the spoken or in written news followed by verbs, adjectives, and adverbs. As the lexical items are those which influence the density of the texts, the discussion is concerned with them. Taking one text from spoken and one text from written news, the nouns which occur are analyzed below.

Nouns in Spoken News

How the Taliban Stormed across Afghanistan in 10 Days

The Taliban fighters sitting behind a very grand desk in the middle of the presidential palace. Taliban fighters have taken down the flag of Afghanistan.

Some districts already been falling to the Taliban hands. Living here in scary and so in hurry, what to do? There's just some of them are running, some of them are keeping themselves hiding in the houses.

The truth is this did unfold more quickly than we had anticipated. So what's happened? Afghanistan political leaders gave up, and fled the country the Afghan military collapsed sometime without trying to fight.

(<https://www.youtube.com/watch?v=xJivUISR6As>)

The text above consists of 104 words and has lexical density level of 50.96%. The dominant content word is a noun with a total of 22 words. The nouns may be classified as common noun (*fighters, district, houses, leaders, flag, country*, etc), proper noun (*Taliban, Afghanistan*), and abstract noun (*hurry, truth*). In other words, common nouns occur more frequently than the other noun. As the topic is Taliban and Afghanistan, the proper nouns which occur refer to them. Meanwhile, another lexical item – verbs – occur 16 times. They are *sitting, taken, falling, living, do, running, keeping, hiding, unforld, anticipated, happened, gave up, fled, collapsed, trying, and fight*. The adjectives which occur in the text are only three items. They are *grand, scary, and political*. *Grand* serves to modify the noun *desk* while *political* modifies the noun *leaders*. The last lexical item is adverbs. They are *already* and *quickly*. The former belongs to adverb of time indicating that something has been done, and the latter is adverb of manner to inform how something is done.

Nouns in Written News

Life under Taliban Rule One Month on

Crossing into the Islamic emirate, this is the border with Uzbekistan. Hundreds of millions of dollars worth of trade pass through here each year. That revenue now is in the hands of the Taliban. For some, life has become much easier.

But sources told us, trade levels have dropped with difficulties making international transactions and some businessmen leaving the country.

An *hour's drive* away the *city* of *Mazzari Sharif*. On the *surface life* appears to be *continuing* as *normal* though many *complain* of *financial pressures*. I want to *see* the *historic Blue Mosque*, the *cultural heart* of the *city*. This was it just a few *days* before the *Taliban takeover* last *month*. Now the *group* have *allocated separate times* for *men* and *women* to visit. Some are *still coming*, but they *seem fewer* than before.

We *literally* just *came* out of the *Blue Mosque* and *saw* a *crowd gathered* in the *city center*. We *made our way* to the *middle* of it, and there are *four dead bodies laid* out here. One of them *has* a *note on top* of it *saying* 'these *men were kidnappers*, *anyone* who *wants* to do the same this is going to *be their punishment*'. All around me there's a *huge crowd* of *people trying* to *push* their way forward to *have a look* at the *site*.

A *group* of *young children* were *rescued* by the *Taliban* from the *kidnappers*. Many *praised* the *group* for *tackling violent crime* that had *plagued major cities*.

But many *others* here don't *feel safe*. Private *universities* like this *one* have *reopened*. *Female* and *male students* are *separated* by a *curtain* as for *new Taliban rules*. But with *money tight* and the *future unclear* only a *handful* are *turning up*.

The *last time* the *Taliban* were in *power*, they *imposed* even more *restrictions*. But however, they may have *evolved*. Since then *Afghan cities* have *changed* much more. The *Taliban* *controlled* the *country*, but *still need* to *win hearts* and *minds*.

(<https://www.bbc.com/news/world-asia-58550640>)

The text above is a sample of written news. It consists of 343 words and has lexical density level of 53.06%. The dominant content word is nouns with a total 90 of words. The noun may be classified as a common noun (*month, sources, businessmen, country, city, month, group, men, women, etc*), proper noun (*Taliban, Uzbekistan, Mazzari Sharif, Blue, and Mosque*), abstract noun (*power, future, and worth*), and collective noun (*A group of young children and a huge crowd of people*). In other words, the same as the first sample, common nouns occur more frequently than the other nouns. Verbs as another lexical item which influence the density of the text appear to be 44 occurrences. Some of the verbs are *crossing, pass, become, told, dropped, making, leaving, appear, continuing, complain, want, see, was, allocated, coming, and seem*. There are 22 adjectives found in the text, such as *Islamic, international, normal, financial, historic, cultural, and unclear*. While, adverbs, like in the spoken news are only found three items (*still, literally*). Based on the discussion above, it is implied that nouns as the dominant lexical item found in both type samples strongly influence the density of the texts.

IV. Conclusion

The present study is about lexical density of spoken and written news of BBC news. Calculating the lexical density using Ure's formula, the findings revealed that both spoken and written news have lexical density higher than 50%. The lexical density of spoken news does not correspond with what Ure's stated which is under 40%. Besides, the lexical density of spoken and written news of BBC News is not as high as what was found by Khamahani [5] in Azeri News and Tehran Times headlines. The spoken and written news of BBC News dense are noun, verb, adjective, and adverb with the dominant lexical item is noun. And there is almost no differences between lexical density of spoken and written news.

Acknowledgment

This paper is based on final project of Wulandari Zahara Handani under the title "Analysis of Lexical Density in Spoken and Written News of BBC News. The final project is the requirement for completing Diploma III Study Program of the English Language in The Department of English Politeknik Negeri Bandung.

References

- [1] E. Castello, *Text complexity and reading comprehension tests*. books.google.com, 2008.

- [2] D. Biber and S. Conrad, *Register, genre, and style: Second edition*. 2019.
- [3] M. A. K. Halliday, *Spoken and written language*. 1985.
- [4] V. Johansson, "Lexical diversity and Lexical Density In Speech and Writing: A Developmental Perspective," *Work. Pap. Linguist.*, vol. 53, pp. 61–79, 2008, [Online]. Available: <https://journals.lub.lu.se/LWPL/article/download/2273/1848>.
- [5] G. Khamahani, "A corpus-based analysis of Tehran Times and Azeri News headlines: Focus on lexical density and readability," *International Journal of Humanities Social Sciences* Citeseer, 2015.
- [6] P. Peters and J. L. Kruger, "The readability of online health information for L1 and L2 Australians: text-based and user-focused research," *Text Talk*, 2021, doi: 10.1515/text-2020-0041.
- [7] V. To, S. Fan, and D. Thomas, "Lexical density and readability: A case study of English textbooks," *Internet Journal of Language, Culture* researchgate.net, 2013, [Online]. Available: https://www.researchgate.net/profile/Vinh-To/publication/257527689_Lexical_density_and_Readability_A_case_study_of_English_Textbooks/links/5c193422a6fdccfc70572a6e/Lexical-density-and-Readability-A-case-study-of-English-Textbooks.pdf.
- [8] M. A. B. Amer, "Lexical Density and Readability of Secondary Stage English Textbooks in Jordan," *International Journal for Management and Modern* researchgate.net, 2021, [Online]. Available: https://www.researchgate.net/profile/Mohammad-Bani-Amer/publication/353247572_Lexical_Density_and_Readability_of_Secondary_Stage_English_Textbooks_in_Jordan/links/60ef5c2b16f9f3130083d34b/Lexical-Density-and-Readability-of-Secondary-Stage-English-Textbook.
- [9] N. Miskiyah, *Analyzing Lexical Density and Readability of Reading Texts in English Textbook Stop Bullying Now by Mahrukh Bashir*. repository.iainkudus.ac.id, 2021.
- [10] H. Syarif and R. E. Putri, "HOW LEXICAL DENSITY REVEALS STUDENTS' ABILITY IN WRITING ACADEMIC TEXT BAGAIMANA DENSITAS LEKSIKAL MENGUNGKAP KEMAMPUAN ...," *download.garuda.kemdikbud.go.id*. [Online]. Available: <http://download.garuda.kemdikbud.go.id/article.php?article=1364341&val=1476&title=HOW LEXICAL DENSITY REVEALS STUDENTS ABILITY IN WRITING ACADEMIC TEXT>.
- [11] M. Ekström, "Epistemologies of TV journalism: A theoretical framework," *Journalism*, 2002, doi: 10.1177/146488490200300301.
- [12] R. Fowler, *Language in the News*. Routledge London, 1991.
- [13] Q. Xie, "Critical discourse analysis of news discourse," *Theory and Practice in Language Studies*. academypublication.com, 2018, [Online]. Available: <https://www.academypublication.com/issues2/tpls/vol08/04/tpls0804.pdf#page=37>.
- [14] A. Radford, *English Syntax*, First. Cambridge: Cambridge University Press, 2004.