



## A Corpus-based Analysis of Tehran Times and New York Times Headlines: Focus on Lexical Density and Readability

Ghaffar Khamahani, Prof. Ilham M. Tahirov

Nesimi Institute of Linguistics, Azerbaijan National Academy of Sciences, Azerbaijan Republic  
[dr-khamahani@yahoo.com](mailto:dr-khamahani@yahoo.com)

**Abstract:** To carry out this research, the total number of headlines arrived at 200 headlines selected from Tehran Times and New York Times. A careful and quantitative analysis of newspaper headlines as a text written by native and non-native speakers was conducted in terms of lexical density. With a careful analysis of lexical density in the two sets of newspaper headlines, the results revealed that New York Times headlines featured a higher lexical density index in comparison to Tehran Times. By taking Flesch's Reading Ease Scale into consideration, this study suggests that headlines written by Tehran Times and New York Times are actually appropriate for native and non-native English students at the postgraduate level, when they are assessed with regards to the American educational attainment levels.

[Ghaffar Khamahani, Ilham M. Tahirov. **A Corpus-based Analysis of Tehran Times and New York Times Headlines: Focus on Lexical Density and Readability**. *N Y Sci J* 2020;13(9):21-25]. ISSN 1554-0200 (print); ISSN 2375-723X (online). <http://www.sciencepub.net/newyork>. 5. doi: [10.7537/marsnys130920.05](https://doi.org/10.7537/marsnys130920.05).

**Keywords:** Lexical Density, Premodification, Headline, Reading Ease

### 1. Introduction

The headline is a unique type of text. It has a range of functions that specifically dictate its shape, content and structure, and it operates within a range of restrictions that limit the freedom of the writer (Reah, 1949). This mix of factors immediately presents a problem: headlines can often be confusing and sometimes ambiguous (Reah, 1949). This problem results from unusual information density. Martin (1993) proposes a formula ( $Inf=Aut/w+c$ ) for calculating what he refers to as "information density". Information density (Inf) is here determined by dividing the number of automatic words (Aut-words which are meaningful on their own, without depending on the context; nouns, main verbs, adjectives and adverbs) in a sentence with number of all words (w) and connectors (c). The purpose of this paper is to investigate lexical density in the headlines of a specific section (World) produced by Tehran Times and New York Times editorial boards. Pedagogical implications are given based on the relationship between lexical density and readability. Several studies of texts written by non-native speakers have been conducted elsewhere (Martin 1993; Chaudron, 2003) as a large research project on the compositions of EFL students. The results of these studies have shown some recurrent patterns. Chaudron (2003) arrived at the conclusion that EFL students use few number of words per clause. In addition, the analysis of noun group (Ure, 1971) shows the use of simple nominal structures with a tendency to post-

modification when communicative needs requires it. Inspired by the previously mentioned studies, this paper aims at a careful and quantitative analysis of newspaper headlines as a text written by native and non-native speakers in terms of lexical density which is the result of premodification. Pedagogical implications regarding the methodology of journalistic texts will be given. Anderson (1975) believes that the rule of *premodification*, the most important factor of lexical density, has been developed so that longer strings of words and phrases now precede the head word in some styles of use to increase lexical density. This style is a particular feature of newspaper *headlines*. Greenbaum & Quirk (1992) affirm that premodification comprises all the items placed before the head -notably determiners, adjectives and nouns. But, concretely, there are more types of premodifying items: we use a fixed frame (his... cottage) - with nonrestrictive function- and we will obtain these series of premodifying items: **a-** Adjective (his delightful cottage) **b-** Participle (his crumbling cottage) **c-** -s genitive (the fisherman's cottage) **d-** noun (his country cottage) **e-** Adverb (his far-away cottage) **f-** Sentence (his pop-down-for-the-weekend cottage).

It is supposed that non-native and native journalists use different types of premodifying items and sometimes multiple premodifying items. This study focuses on statistically quantifying lexical density produced by premodification and sometimes

postmodification items included in the two sets of the headlines of the newspapers, namely Tehran Times and New York Times.

### Objectives of the Study

This study is mainly concerned with: **a-** arriving at the exact amount of lexical density in Tehran Times and New York Times headlines **b-** focusing on syntactic differences between the two sets of data **c-** giving pedagogical implications for teaching journalistic courses.

### Review of the Related Literature

**Lexical Density, Premodification and Postmodification:** In computational linguistics, lexical density constitutes the estimated measure of content words per functional (grammatical) words and lexical units (lexemes) (WIKIPODIA, The Free Encyclopedia). Halliday (1989) defines lexical density as '... a measure of the density of information in any passage of text, according to how tightly the lexical items (content words) have been packed into the grammatical structure. It can be measured, in English, as the number of lexical words per clause. Ure (1971) defines lexical density as a ratio of the number of lexical words per clause. According to Thompson (1996) lexical density is related, in English, to the use of noun group structure. Apart from the clauses, the nominal group is "the grammatical unit" which allows the widest range of meaning to be expressed. Nominal groups usually carry over the lexical density of texts, since they pack information within their possibly extended structure of premodification- head-postmodification or qualifier (Halliday 1994). He believes that the potential meaning expression of the nominal group is identified with the construction of highly informative texts, prototypically associated with written style. In those kinds of texts, complex relationships are expressed "nominally" rather than "clausally" as in spoken style. Coffin (1997) contends that from a functional perspective written discourse is dominated by a nominal style. Like all phrases, the constituents of the English noun phrase can be analyzed into both functional constituents and formal constituents. From a functional point of view, the noun phrase has four major components, occurring in a fixed order:

- the *determinative*, that constituent which determines the reference of the noun phrase in its linguistic or situational context;
- *premodification*, which comprises all the modifying or describing constituents before the head, other than the determiners;
- the *head*, around which the other constituents cluster; and
- *Postmodification*, those which comprise all the modifying constituents placed after the head.

According to the preceding literature, it seems necessary to elaborate on content words. Content words are words that have meaning. They can be compared to grammatical words, which are structural. Nouns, main verbs, adjectives and adverbs are usually content words. Auxiliary verbs, pronouns, articles, and prepositions are usually grammatical words. For example in the sentence 'We flew over the mountains at dawn' the underlined words are content words, but the remaining words are functional.

A prepositional phrase is made up of a preposition and a noun phrase. We use prepositional phrases for many purposes, for example:- as adverbials of time and place: We will be back in a few days. They drove to Glasgow.- as a postmodifier in a noun phrase: Helen is the girl in the red dress. We've got a new television with a thirty one inch screen.- to show who did something: The lion was killed by the hunter. I saw a wonderful painting by Van Gog- with double object verbs like give and get: We gave five pounds to the woman on the corner. They got a drink for me.- after certain verbs, nouns and adjectives: The book belongs to me. I had an argument with my brother. I feel sorry for you.

## 2. Research Methodology

**Data Collection and Analysis:** Some analyses concerning grammatical premodification, including the use of deictics, post deictics and numeratives regarding writing style have been carried out. This study focuses on lexical density concerning newspaper headlines, as a type of written text. Adopting a nonrandom purpose sampling, the present study selected 100 headlines culled from the electronic version of Tehran Times and New York Times English newspapers over a specific time period (April 25th-April 30th). As it was stated earlier, the headlines were selected from only one section of the newspapers (World). In order to carry out the research, the total number of words including content words and functional words in the newspaper headlines were put into word format and analyzed by Lexical Density Test to arrive at the ratio of content words to functional words.

The Lexical Density Test is another readability test that is designed to measure the degree of reading difficulty that a text presents. The Lexical (content) density is calculated by the following formula: Lexical Density = (number of different words) ÷ (total number of words) × 100.

The lexical density of a text is an attempt to measure the proportion of the content (lexical) words over the total words. Texts with a lower density are more easily understood. A score of 60-70% indicates a text that is lexically dense. A score of 40-50% is indicative of text that is not considered to be dense.

Table 1: Tabulates Total Number of Words, Total Number of Content Words, Total Number of Functional Words, Lexica Density (Percentage) in Tehran Times

Total Number of Words	Total Number of Content Words	Total Number of Functional Words	Lexical Density
980	738	242	% 75

Table 2: Tabulates Total Number of Words, Total Number of Content Words, Total Number of Functional Words, Lexica Density (Percentage) in New York Times.

Total Number of Words	Total Number of Content Words	Total Number of Functional Words	Lexical Density
830	737	93	%88

Table 3: Tabulates the Total No. of Headlines , Nominal Groups (Premodified) , Nominal Groups (Postmodified) , Nouns Lacking Pre- or Postmodification items , Prepositional Phrases and Nominal Groups in Prepositional Phrases In Tehran Times

No. of Headlines	No. of Nominal Groups(Premodified)	No. of Nominal Groups (Postmodified)	Nouns Lacking Pre- or Postmodification	No. of Prepositional Phrases	No. of Nominal Groups in Prepositional Phrases
100	310	15	180	270	270

Table 4: Tabulates the Total No. of Headlines , Nominal Groups(Premodified) , Nominal Groups (Postmodified) , Nouns Lacking Pre- or Postmodification Items , Prepositional Phrases and Nominal Groups in Prepositional Phrases in New York Times.

No. of Headlines	No. of Nominal Groups(Premodified)	No. of Nominal Groups (Postmodified)	Nouns Lacking Pre- or Postmodification	No. of Prepositional Phrases	No. of Nominal Groups in Prepositional Phrases
100	360	25	125	260	260

With a careful analysis of lexical density in the two sets of newspaper headlines, the results reveal that New York Times headlines featured a higher lexical density index in comparison to Tehran Times. As far as Table 1 and Table 2 are concerned, the dominant feature increasing lexical density in New York Times is the remarkable use of content words by New York Times headline writers. However the two sets of headlines have a high lexical density when considered with Halliday (1985) and Ure's (1971) measurements, and are rated as very difficult in terms of Flesch's Reading Ease Scale.

As stated earlier, according to Thompson (1991) lexical density is related, in English, to the use of noun group structure. Apart from the clauses; the nominal group is "the grammatical unit" which allows the widest range of meaning to be expressed. Halliday

(1994) contends that nominal groups usually carry over the lexical density of texts, since they pack information within their possibly extended structure of premodification- head- postmodification or qualifier. By taking Thompson's and Halliday's perspectives into consideration, we arrive at this major result. The most lexical density was observed in nominal groups, which were mainly premodified in both sets of headlines.

By using Online Text analysis Tool, the exact lexical density was obtained. The following table shows the ratio of content words to functional words in Tehran Times and New York Times newspapers.

In order to arrive at statistically quantitative results, the headlines were analyzed in terms of premodified nominal groups (either premodified by a noun (s) or adjective (s)), nominal groups which were

postmodified, nouns lacking premodification or postmodification items, prepositional phrases and nominal groups in prepositional phrases. As far as Tables 3 and Table 4 concern, premodified nominal groups have topmost position in terms of lexical index. The results of this study suggest that New York Times uses more postmodified nominal groups. Nouns without premodifying or postmodifying items were much more frequently observed in New York Times. These factors ultimately results in a higher lexical index in New York Times headlines. The frequency of the occurrence of noun phrases found in prepositional phrases was more or less the same in the two sets of headlines.

### Conclusion and Pedagogical Implications:

In order to give pedagogical implications based on the results, we employed methods proposed by Ure (1971) and Halliday (1980), which are central measurements to lexical density exploration in texts (in this case headlines): Formula 1 (Ure's method):

$$\text{Lexical Density} = \frac{\text{Number of Lexical Items} * 100}{\text{Total Number of Words}}$$

(Halliday,1985; O'loughlin,1995; Ure,1971)

According to this formula, if the amount of lexical density surpasses 50 percent, it indicates a high lexical density. By taking Flesch's Reading Ease scale into consideration, the higher the density the more complicated the text is. Since this study intends to give pedagogical implications using this method, an outline of the method concerning readability is given below:

### Flesch's Reading Ease Scale

By taking this scale into consideration, this study suggests that headlines written by Tehran Times and New York Times are actually appropriate for native and non-native English students at the postgraduate level, when they are assessed with regards to the American educational attainment levels.

Flesch Reading Ease	Description Of Style	Educational Attainment Level (U.S.A)
0-10	Very easy	Grade 5
10-20	Easy	Grade 6
20-30	Fairly easy	Grade7
30-40	Standard	Grade8-9
40-50	Fairly difficult	Grade10-12
50-70	difficult	Undergraduate
70-100	Very difficult	Postgraduate

### Corresponding Author:

Ghaffar Khamahani

Nesimi Institute of Linguistics, Azerbaijan National Academy of Sciences, Azerbaijan Republic.

E-mail: [dr-khamahani@yahoo.com](mailto:dr-khamahani@yahoo.com)

### References:

- MARTIN, J. R. 1993 The discourse of Geography: ordering and explaining the experiential world. IN: M. A. K., HALLIDAY, & J. R. MARTIN (Eds). Writing Science: Literacy and the Discourse of Power. The Falmer Press. 133-165.
- Halliday, M. A. K. and Martin, J. R. 1993. Writing Science: Literacy and Discursive Power. Pittsburgh: University of Pittsburgh Press. pp. 4-6, 8-11.
- Ure, J. (1971). Lexical density and register differentiation. Applications of linguistics, 443-452.
- O'Loughlin, K. (1995). Lexical density in Candidate output on two versions of An oral Proficiency Test. Melbourne Papers in Language Teaching, 26-48.
- Andersson L. G., Form and Function of Subordinate Clauses, 1975, Oxford, England: Blackwell.
- Greenbaum, S & R. Quirk. (1992, p. 346) A student's Grammar of the English Language, Essex: Longman.
- Chaudron, C (2003) Data Collection in SLA Research. In C. Doughty & M. long (Eds.), The handbook of second language acquisition (pp.762-858). Oxford, England; Blackwell.
- Reah D., 1949, The language of newspapers, 2 nd ed., London, Routledge 2002.
- Halliday M. A. (1989) Spoken and Written Language London: Longman.
- Halliday M. A. (1994) Introduction to Systematic Functional Linguistics
- Coffin G (1997) cited in Ken Hyland (2002). 6. GENRE: LANGUAGE, CONTEXT, AND LITERACY. Annual Review of Applied

- Linguistics, 22, pp 113-135.  
doi:10.1017/S0267190502000065
12. Bourdieu, P., & Thompson, J. B. (1991). Language and symbolic power. Harvard University Press.
  13. Flesch, R. F. (1948). "A New Readability Yardstick," Journal of Applied Psychology, 32, 221-223.
  14. **On-line References**  
www.tehrantimes.com
  15. www.nytimes.com
  16. [http://papyr.com/hypertextbooks/grammar/ph\\_no\\_un.htm](http://papyr.com/hypertextbooks/grammar/ph_no_un.htm)
  17. <http://www.teachingenglish.org.uk/knowledge-database/content-words>
  18. <http://learnenglish.britishcouncil.org/en/english-grammar/clause-phrase-and-sentence/prepositional-phrases>.
  19. <http://www.englishproofreading.ca/lexical.htm>
  20. <http://aaref.com.au/en/publications/journal>

9/24/2020