

ABSTRACT

Herdian, Damasus Desta. (2017). *A Corpus Analysis of Vocabulary Coverage and Word Frequency of Junior High School Course Books*. Yogyakarta: Sanata Dharma University

One of the important keys in learning English is through acquiring vocabulary from the course books. One of the ways to validate vocabulary coverage and word frequency in the course book is by comparing the vocabulary with corpus GSL (General Service List). The corpus words were selected to represent the most frequent words of English and were taken from a corpus of written English.

This research was aimed to discover the characteristics of the vocabulary coverage and word frequency in Junior High School course books, grade Eight. With two research questions, which are: (1) What is the vocabulary coverage of Junior High School course books in terms of types, tokens, and word families? (2) How frequent the words are being used in the course books?

The research method was corpus-based survey. The data were obtained from four junior high school course books, published by official/government and private publishers. Two Paul Nation's concordance programs and a modified program (based on Bauman and Culligan's wordlists) were used as instruments for data analysis. They were RANGE and FREQUENCY.

The findings showed that the four course books have average 30,463 tokens, and 3,063 types. The course books covered 72.55% tokens from total first 800 words (GSL800) vocabulary which should be mastered as Junior High School learners. The course books covered average 622 out of 800 word families or 77.78% of the total 800 headwords (GSL800) as required for the 2nd grade of Junior High School learners. It is considered not enough and need some improvements. The contents of the course books are accessible enough to students whose vocabulary knowledge is within the range. Unfortunately, they have inadequate number of word families and students will find difficulties in producing speaking and writing (active English) unless they use several supplementary English books or worksheets. The course books do not meet the criteria of incidental learning and text coverage either the learning opportunity to the students is not really high. The average deficit vocabulary is 177 out of 800 headwords. Almost 54% word families from the course books are repeated less than ten times. Thus, it gives small opportunities to students for deepening learning vocabulary. There are 20 functional words and 4 content words in the total 24 most frequent word types from all course books. The words *the, to, a, and* are four most frequent words appeared in all course books.

The course books have most of the types, tokens, and families listed in the GSL wordlist. However, authors need significant improvement in designing the material to give students a higher opportunity in learning vocabulary. High amount of tokens, types, and word families are also necessary but should be followed with suitable vocabulary needs and high frequency of essential words to help students become more effective and efficient in learning vocabulary.

Keywords: vocabulary coverage, word frequency, course book, junior high school

ABSTRAK

Herdian, Damasus Desta. (2017). *A Corpus Analysis of Vocabulary Coverage and Word Frequency of Junior High School Course Books*. Yogyakarta: Sanata Dharma University

Salah satu kunci utama dalam belajar Bahasa Inggris adalah penguasaan kosakata lewat buku materi. Salah satu cara untuk validasi cakupan kosakata dan pengulangannya adalah dengan membandingkannya terhadap korpus GSL. Korpus ini mewakili kosakata bahasa Inggris yang paling sering muncul yang diambil dari banyak sumber berbahasa Inggris.

Penelitian ini bertujuan untuk menemukan karakteristik kosakata dan pengulangan kata di beberapa buku materi SMP kelas delapan. Ada dua permasalahan yang dipecahkan di penelitian ini: (1) Cakupan kosakata di buku pegangan SMP, menemukan jumlah *token*, *type*, dan *word family* (2) Bagaimana kosakata digunakan ulang, untuk menjelaskan bagaimana kata digunakan ulang.

Metode penelitian ini berbasis korpus. Korpus didapatkan dari empat buku materi untuk SMP yang diterbitkan oleh pemerintah maupun penerbit swasta. Dua buah program konkordansi dari Paul Nation dan program modifikasi (berdasar daftar kata dari Bauman dan Culligan) digunakan sebagai instrumen analisis data. Program itu bernama RANGE dan FREQUENCY Hasil analisis data diinterpretasi untuk menemukan jawaban atas pertanyaan penelitian ini.

Hasil penemuan menunjukkan bahwa empat buku tersebut memiliki rata-rata 30.463 *tokens*, dan 5.063 *types*. Ini mencakup 72,55% *tokens* dari total kosakata GSL800 yang harus dikuasai siswa SMP. Buku mencakup 622 dari 800 *word families* atau 77,78% dari (GSL800) yang diharapkan siswa SMP mencapainya (daftar kata Bauman dan Culligan). Ini dikategorikan rendah dan perlu peningkatan. Buku mudah dipelajari oleh siswa yang memiliki pengetahuan kosakata sesuai levelnya. Sayangnya sedikitnya varian *types* dan *word families* membuat siswa sulit mempelajari bahasa Inggris secara aktif, kecuali ditambah buku penunjang atau buku soal. Buku materi juga tidak menunjang siswa untuk belajar mandiri diluar KBM. Kesempatan untuk mempelajari kosakata yang lebih sulit tidak banyak diberikan. Rata-rata kosakata yang kurang adalah 177 dari 800 kata. Hampir sekitar 54% *word families* dari keseluruhan buku materi digunakan kurang dari 10 kali. Artinya, kecil kemungkinan bagi siswa untuk lebih mendalami kosakata dengan baik. Ada 20 *functional words* dan 4 *content words* dari total 24 kata yang paling sering muncul dari seluruh buku materi. Kata *the*, *to*, *a*, dan *and* adalah empat *functional words* yang paling sering muncul.

Dapat disimpulkan bahwa buku materi memiliki sebagian besar *tokens*, *types*, dan *word families* yang terdapat pada daftar kata bahasa Inggris yang sering muncul. Namun, buku tetap membutuhkan peningkatan yang signifikan dalam desain material untuk memberikan kepada siswa sebanyak mungkin kesempatan belajar kosakata. Besarnya *tokens*, *types*, dan *word families* itu perlu, namun harus diikuti oleh kebutuhan kosakata yang sesuai dan pengulangan kosakata untuk membantu siswa lebih efektif dan efisien dalam mempelajari kosakata.

Kata kunci: cakupan kosakata, frekuensi kata, buku materi, SMP