

# Nghiên cứu các tổ hợp từ trong bộ sách giáo khoa Tiếng Anh cho học sinh phổ thông ở Việt Nam

Cao Thị Phương Dung\*, Phó Phương Dung, Đặng Nguyễn Anh Chi



Use your smartphone to scan this QR code and download this article

## TÓM TẮT

Tổ hợp từ (collocations) đóng vai trò rất quan trọng trong việc dạy và học ngoại ngữ. Trong môi trường mà tiếng Anh được dạy như là một ngoại ngữ thì sách giáo khoa là nguồn học liệu chính mà người học có thể tiếp cận. Do vậy, bài viết này nghiên cứu về việc sử dụng tổ hợp từ trong bộ sách giáo khoa tiếng Anh dành cho học sinh các cấp tiểu học, trung học cơ sở và trung học phổ thông ở Việt Nam, tập trung vào việc nghiên cứu (1) sự tương thích của các tổ hợp từ được chọn là trọng tâm trong bộ sách với danh mục các tổ hợp từ có tần suất xuất hiện cao cần được giới thiệu cho người học, và (2) việc tái sử dụng các tổ hợp từ được chọn là trọng tâm này trong bộ sách. Từ bộ sách giáo khoa, nhóm nghiên cứu xây dựng được kho ngữ liệu gồm 312.770 từ, trong đó có 13.292 tổ hợp từ có cấu trúc động-danh và 11.079 tổ hợp từ có cấu trúc tính-danh. Tần suất xuất hiện của các hạng mục tổ hợp từ và dạng thức tổ hợp từ tăng dần theo cấp lớp. Kết quả nghiên cứu cho thấy các tổ hợp từ được chọn là trọng tâm trong bộ sách chỉ chiếm khoảng 10,5% danh mục các tổ hợp từ tiếng Anh học thuật. 11,4% tổ hợp từ trọng tâm được giới thiệu lặp lại ở các đơn vị bài học khác nhau và 19,3% được xác định không phải là tổ hợp từ có tần suất xuất hiện cao. Ngoài ra, có đến hơn ba phần tư tổ hợp từ trọng tâm không được tái sử dụng hoặc việc tái sử dụng chưa đạt đến tần suất nhất định để có thể giúp người học tiếp thu và sẵn sàng cho việc sử dụng. Dựa trên các kết quả tìm được, bài nghiên cứu bàn luận về việc dạy và học cũng như biên soạn tài liệu liên quan đến các tổ hợp từ.

**Từ khoá:** tổ hợp từ, sách giáo khoa dành cho người học tiếng Anh như một ngoại ngữ, danh mục các tổ hợp từ, tiếp thu từ vựng

## GIỚI THIỆU

Tổ hợp từ đóng vai trò rất quan trọng trong việc dạy và học ngoại ngữ. Những nghiên cứu xoay quanh lĩnh vực này cho thấy việc dạy và học các tổ hợp từ giúp cải thiện vốn từ vựng và khả năng giao tiếp của người học. Biết dùng tổ hợp từ sẽ giúp cho người học tự tin hơn trong giao tiếp và sử dụng ngoại ngữ giống với cách người bản ngữ sử dụng hơn. Tuy nhiên, những nghiên cứu trước đây cũng cho thấy không dễ dàng để nâng cao kiến thức về tổ hợp từ, và nhiều người học ngoại ngữ dù là ở cấp độ chuyên sâu vẫn gặp khó khăn với việc sử dụng một cách chính xác các tổ hợp từ, đặc biệt là những tổ hợp từ có cấu trúc động-danh và tính-danh<sup>1-3</sup>.

Trong môi trường mà tiếng Anh được xem như là một ngoại ngữ, tức là người học khó có thể đạt được độ lưu loát và tự nhiên như người bản ngữ, thì sách giáo khoa trở thành một trong những nguồn chính để dung nạp ngôn ngữ đích cho người học. Ở Việt Nam, bộ sách giáo khoa dành cho các cấp tiểu học, trung học cơ sở và trung học phổ thông được các nhóm tác giả biên soạn và lưu hành bởi Bộ Giáo dục và Đào tạo sau khi được đánh giá chất lượng. Theo thông tư 32/2018 của

Bộ Giáo dục và Đào tạo, một trong các tiêu chí xây dựng bộ sách giáo khoa là nhằm cung cấp cho người học lượng từ vựng cần thiết cho việc giao tiếp cơ bản. Nội dung dạy học Chương trình giáo dục phổ thông tiếng Anh hầu hết được thiết kế theo kết cấu thành phần bao gồm (1) hệ thống các chủ điểm, chủ đề, (2) năng lực giao tiếp liên quan đến các chủ điểm, chủ đề, (3) danh mục các kiến thức ngôn ngữ (phát âm, từ vựng, ngữ pháp). Nội dung từ vựng ở cấp tiểu học là những từ thông dụng, đơn giản ở Bậc 1 trong *Khung năng lực ngoại ngữ sáu bậc* nhằm phục vụ cho các tình huống giao tiếp đơn giản theo chủ điểm, chủ đề của chương trình. Số lượng từ vựng bộ sách giáo khoa cần cung cấp cho học sinh ở cấp tiểu học là khoảng 600-700 từ. Nội dung từ vựng ở cấp trung học cơ sở bao gồm những từ thông dụng trong lĩnh vực ngôn ngữ nói và ngôn ngữ viết liên quan các chủ điểm, chủ đề trong chương trình. Ở cấp học này bộ sách cần cung cấp thêm cho người học khoảng 800-1.000 từ ở Bậc 2. Nội dung từ vựng ở cấp trung học phổ thông cũng gồm những từ thông dụng trong cả hai lĩnh vực ngôn ngữ nói và viết. Số lượng từ vựng được quy định ở cấp học này khoảng 600-800 từ ở Bậc 3. Khi hoàn

Trường Đại học Khoa học Xã hội và Nhân văn, ĐHQG-HCM, Việt Nam

### Liên hệ

Cao Thị Phương Dung, Trường Đại học Khoa học Xã hội và Nhân văn, ĐHQG-HCM, Việt Nam

Email: ctpdung@hcmussh.edu.vn

### Lịch sử

- Ngày nhận: 28-11-2021
- Ngày chấp nhận: 20-5-2022
- Ngày đăng: 30-6-2022

DOI: 10.32508/stdjssh.v6i2.723



### Bản quyền

© ĐHQG Tp.HCM. Đây là bài báo công bố mở được phát hành theo các điều khoản của the Creative Commons Attribution 4.0 International license.



**Trích dẫn bài báo này:** Cao, T. P. D., Phó, P. D., & Đặng, N. A. C. **Nghiên cứu các tổ hợp từ trong bộ sách giáo khoa Tiếng Anh cho học sinh phổ thông ở Việt Nam.** *Sci. Tech. Dev. J. - Soc. Sci. Hum.*; 2022 6(2):1573-1586.

tất cấp học trung học phổ thông, học sinh sẽ lĩnh hội được khoảng từ 2.500 đến 2.800 từ vựng tiếng Anh, tương ứng với danh mục Từ vựng phổ thông (NGSL) gồm 2.801 từ do Browne và các cộng sự<sup>4</sup> xây dựng dựa trên kho ngữ liệu Cambridge English 1,6 tỷ từ. Nhận thức được tầm quan trọng của tổ hợp từ, các nhà biên soạn sách giáo khoa tiếng Anh thường có khuynh hướng đưa nội dung này vào mục giới thiệu từ vựng trong mỗi đơn vị bài học. Bộ sách giáo khoa tiếng Anh dành cho các cấp tiểu học, trung học cơ sở và trung học phổ thông Việt Nam cũng không ngoại lệ. Tuy nhiên, chưa có nghiên cứu nào tập trung vào việc xem xét liệu các tổ hợp từ mà bộ sách giáo khoa tiếng Anh giới thiệu đến học sinh có tập trung vào các tổ hợp từ có tần suất xuất hiện cao và hữu ích với người học hay không. Ngoài ra, cũng không rõ là các tổ hợp từ được chọn làm trọng tâm trong bộ sách giáo khoa tiếng Anh có được tái sử dụng một cách hệ thống tạo điều kiện thuận lợi cho việc học hay không. Theo các nghiên cứu về tiếp thu ngôn ngữ, việc chỉ giới thiệu một lần các từ vựng đơn lẻ nói chung và tổ hợp từ nói riêng là chưa đủ để giúp người học tiếp thu và sẵn sàng cho việc sử dụng. Việc tái sử dụng, lặp lại các từ vựng đã dạy sẽ giúp trí nhớ ngắn hạn “chuyển thành trí nhớ dài hạn”<sup>5</sup>. Do đó, bài nghiên cứu này tập trung vào các tổ hợp từ có kết cấu động-danh và tính-danh và nhằm để trả lời cho các câu hỏi nghiên cứu sau đây:

1. Các tổ hợp từ trọng tâm trong bộ sách giáo khoa tiếng Anh này có tương thích với danh mục các tổ hợp từ cần được giới thiệu đến người học hay không?
2. Các tổ hợp từ trọng tâm trong bộ sách giáo khoa tiếng Anh này có được tái sử dụng để tạo điều kiện thuận lợi cho việc tiếp thu tổ hợp từ tiếng Anh hay không?

## TỔNG QUAN

### Nhận diện tổ hợp từ

Các tổ hợp từ được định nghĩa khác nhau tùy theo việc chúng được xem xét từ góc độ tần suất cùng xuất hiện của các từ trong một cụm từ (frequency-based approach) hay ngữ cú của chúng (phraseology-based approach)<sup>2</sup>. Những học giả theo cách tiếp cận thống kê định nghĩa tổ hợp từ là sự cùng xuất hiện của các từ ở một mức độ thường xuyên nào đó<sup>6-9</sup>. Chỉ số “cùng xuất hiện” (MI-score) trong kho ngữ liệu Anh ngữ được sử dụng để đo tần suất cùng xuất hiện của hai hay nhiều hơn các từ một cách ngẫu nhiên<sup>10,11</sup>. Cũng theo Hunston<sup>10</sup> chỉ số này thường được sử dụng trong nhiều nghiên cứu như là một phép đo lường thống kê để nhận diện tổ hợp từ. Chỉ số MI 3,0 do Jones và Sinclair<sup>11</sup> đề xuất đã được sử dụng

ở nhiều nghiên cứu khi tính toán tần suất cùng xuất hiện của các từ<sup>12</sup>. Đối với các học giả theo cách tiếp cận thống kê này, cấu trúc giữa các thành phần trong một cụm từ (động-danh, tính-danh) không đóng vai trò quyết định liệu cụm từ này có phải là một tổ hợp từ hay không<sup>13</sup>. Tuy nhiên, chỉ số MI-score không hẳn là một phép thống kê hiệu quả bởi vì nó nhấn mạnh vào những cặp từ hiếm gặp<sup>14-16</sup>. Ví dụ hai từ *prophecy* và *disaster* cùng xuất hiện liền kề nhau chỉ có một lần trong Kho ngữ liệu quốc gia Anh ngữ (BNC) nhưng lại có chỉ số MI-score rất cao (8,21)<sup>17</sup>. Thấy được điểm yếu này của chỉ số MI, một số nhà nghiên cứu dùng thêm chỉ số t-score (> 2,0) như là một tiêu chuẩn bổ sung trong việc nhận dạng tổ hợp từ để bảo đảm rằng một kết hợp từ phải xuất hiện đủ thường xuyên trong quá trình sử dụng ngôn ngữ mới được xem là một tổ hợp từ<sup>17</sup>.

Trái lại, theo các học giả chọn cách tiếp cận ngữ cú, việc xem xét liệu một cụm từ có phải là tổ hợp từ hay không phải xét đến kết cấu ngữ pháp trước nhất<sup>18-20</sup>. Một tổ hợp từ luôn gồm có một từ gốc và từ liên kết. Từ gốc sẽ là danh từ trong những kết cấu có chứa danh từ như động-danh (*take responsibility*) và tính-danh (*natural resources*). Cách nhận diện tổ hợp từ dưới góc độ này thường phân biệt tổ hợp từ với cách kết hợp từ tự do và thành ngữ<sup>21</sup>. Tính “thay thế có giới hạn” (restricted substitutability)<sup>2</sup> là một tiêu chí được các học giả theo cách tiếp cận này sử dụng rộng rãi để phân biệt giữa tổ hợp từ và cách kết hợp từ tự do<sup>2,19,22</sup>. Đối với cách kết hợp từ tự do thì tính thay thế của các thành phần trong một cụm từ chủ yếu là dựa vào các đặc điểm về ngữ nghĩa của các từ trong cụm từ đó. Sự phân biệt giữa tổ hợp từ và thành ngữ lại dựa trên tính rõ nghĩa của mỗi cụm từ. Có thể xem *miss a chance* và *rain cats and dogs* lần lượt là những ví dụ về tổ hợp từ và thành ngữ bởi vì nghĩa của chuỗi từ trước là sự kết hợp nghĩa của từng từ riêng lẻ trong khi nghĩa của chuỗi từ sau lại không phải vậy. Nói một cách khác, nghĩa của một thành ngữ không phải là sự kết hợp về nghĩa của các từ trong chuỗi từ đó. Theo đó, các tổ hợp từ trong nghiên cứu này sẽ được nhận diện dựa trên sự kết hợp của cả cách tiếp cận thống kê lẫn tiếp cận ngữ cú. Cách kết hợp này cho phép nhóm nghiên cứu tận dụng tối ưu phép đo tần suất của cách tiếp cận thống kê đồng thời chỉ tập trung vào các tổ hợp từ có cấu trúc cụ thể là động-danh và tính-danh. Đây là các cấu trúc mà, như đã đề cập ở trên, người học thường hay mắc lỗi sai. Với cách xác định tổ hợp từ này, những cụm từ có chỉ số MI và chỉ số t-score cao như *and respectively*, *between and*<sup>7</sup> hay *any other*, *each other*<sup>23</sup> vốn không có giá trị cao trong việc dạy và học sẽ không được xem là tổ hợp từ.

### Danh mục các tổ hợp từ

Một số danh mục tổ hợp từ có tần suất xuất hiện cao được các học giả xây dựng với mục đích tập trung vào hoạt động dạy và học. Danh mục tổ hợp từ của Shin và Nation<sup>23</sup> gồm 100 tổ hợp từ tiếng Anh tổng quát có thể nói là danh mục tổ hợp từ được xây dựng đầu tiên. Theo Shin và Nation<sup>23</sup> việc xây dựng danh mục này là nhằm để giới thiệu cho người học tiếng Anh ở trình độ sơ cấp và tiền trung cấp – tức những người mà vốn từ chỉ đầu đó khoảng 1.000 từ. Để xây dựng danh mục này, nhóm tác giả đã dùng kho ngữ liệu BNC 10 triệu từ làm nguồn dữ liệu. Họ xây dựng dựa trên bốn tiêu chí: (1) từ gốc phải là một danh từ, động từ, tính từ, hay trạng từ; (2) chúng phải thuộc nhóm 1.000 từ nội dung (content words) thường xuyên xuất hiện nhất trong kho ngữ liệu BNC; (3) các tổ hợp từ phải xuất hiện ít nhất 30 lần trong kho ngữ liệu; và (4) các tổ hợp từ phải đúng ngữ pháp. Vì chỉ có từ gốc trong tổ hợp từ là từ nội dung nên phần lớn các tổ hợp từ trong danh mục này (ví dụ: *a bit, as well, said to*) không có nghĩa và vì vậy không thực sự hữu ích để giới thiệu đến người học. Nếu xem xét kỹ danh mục này sẽ thấy chỉ có 6 tổ hợp từ có kết cấu động-danh và tính-danh. Như vậy, danh mục tổ hợp từ này không phải là hình mẫu cho công trình nghiên cứu tập trung vào các tổ hợp từ mà cả từ gốc và từ liên kết đều là từ nội dung như nghiên cứu này.

Bên cạnh danh mục tổ hợp từ của Shin và Nation<sup>23</sup>, danh mục tổ hợp từ của Durrant<sup>7</sup> gồm 1.000 từ lại là danh mục đầu tiên tập trung vào các tổ hợp từ học thuật. Tác giả mô tả rằng để xây dựng danh mục, ông đã thu thập một kho ngữ liệu gồm 25 triệu cụm từ từ các bài nghiên cứu trong các tạp chí chuyên ngành. Rồi từ kho ngữ liệu tự thu thập đó, ông lọc ra những tổ hợp từ có tần suất xuất hiện cao thông qua một bộ tiêu chí chọn lọc và những tiến trình chặt chẽ. Cụ thể là Durrant bắt đầu bằng việc trích xuất các chuỗi từ thường xuyên xuất hiện trong những văn bản học thuật hơn là trong những văn bản phi học thuật. Ngoài ra, những cụm từ này phải có tần suất xuất hiện tối thiểu một lần trong kho ngữ liệu học thuật. Một tiêu chí khác phải được đáp ứng là chỉ số MI của những cặp từ này phải trên mức 4,0. Mặc dù đã được xây dựng qua một tiến trình chặt chẽ, danh mục vẫn còn một vài hạn chế. Thứ nhất, nó bao gồm những cụm từ không hoàn chỉnh về mặt cấu trúc (ví dụ: *identified been, effect long, factor such*). Thứ hai, có nhiều cụm từ mà sự kết hợp giữa chúng là hoàn toàn tự do như *these features/ these conditions*. Các cụm từ này không có nhiều giá trị trong việc dạy cho người học. Ngoài ra, còn có một danh mục tổ hợp từ học thuật khác nữa do Chon và Shin<sup>24</sup> xây dựng.

Danh mục này cũng được xây dựng thông qua một tiến trình gồm các bước khá chặt chẽ tuy nhiên cũng không tốt hơn danh mục của Durrant<sup>7</sup> vì những lý do tương tự. Cụ thể, danh mục này cũng bao gồm rất nhiều các cụm từ tự do như *these data/this data* vốn không gây khó khăn nhiều cho người học và vì vậy không có nhiều giá trị trong giảng dạy.

Một danh mục tổ hợp từ học thuật khác nữa đã được xây dựng bởi Ackermann và Chen<sup>25</sup>. Danh mục này được cho là vượt qua những hạn chế của các danh mục đã đề cập bên trên ở chỗ nó chỉ chứa những tổ hợp từ hoàn chỉnh về cấu trúc<sup>26</sup>. Các tổ hợp từ trong danh mục này có các cấu trúc phổ biến mà hầu hết giáo viên cũng như các nhà từ điển học đều tập trung vào (ví dụ: động-danh, tính-danh, trạng-tính, động-trạng). Để xây dựng danh mục này, các tác giả đã áp dụng một tiến trình gồm năm bước: (1) dùng chức năng POS-tagging (gán nhãn từ loại) trong kho ngữ liệu Anh ngữ để trích xuất các cụm từ có cấu trúc cụ thể từ danh mục 1.300 từ nội dung (đây là các từ có tần suất xuất hiện cao trong BNC); (2) lọc ra những tổ hợp từ đáp ứng được các chỉ số đo lường thống kê với MI score  $\geq 3$  and t-score  $\geq 4$ ; (3) kiểm tra theo cách thủ công để loại bỏ các cụm từ không hoàn chỉnh về mặt cấu trúc; (4) đánh giá chuyên môn về tính thích hợp của các cụm từ này cho mục đích giảng dạy; và cuối cùng là (5) hệ thống hoá danh mục (ví dụ bằng cách thêm các mạo từ hay thêm các giới từ). Tiến trình này mang lại một danh mục gồm 2.468 tổ hợp từ trong đó có 1.773 là tính-danh và 310 là động-danh. Tránh được những hạn chế của hai danh mục kia, danh mục do Ackermann và Chen<sup>26</sup> xây dựng trở thành một nguồn ngữ liệu tổ hợp từ học thuật có giá trị cho người học. Vì vậy, nó được dùng như hình mẫu tham chiếu trong nghiên cứu này.

### Các khuyến nghị khoa học về cách trình bày và giảng dạy tổ hợp từ

Ở một nghiên cứu khoa học Nation<sup>27</sup> đã chỉ ra rằng việc dạy từ vựng một cách chủ động có tác động rất quan trọng đến khả năng ghi nhớ dài hạn ở người học. Điều này cũng được chứng minh tương tự đối với tổ hợp từ<sup>18,28,29</sup>. Tuy nhiên, với lượng thời gian eo hẹp trên lớp, việc dạy chủ động một khối lượng lớn từ vựng nói chung và tổ hợp từ nói riêng dường như lại là một nhiệm vụ bất khả thi. Để bù lại cho việc phải dạy chủ động một khối lượng lớn tổ hợp từ với quỹ thời gian hạn hẹp trên lớp, Webb và các cộng sự<sup>30</sup> khuyến nghị việc nâng cao nhận thức của người học về sự hiện diện của tổ hợp từ. Một khi ý thức được sự tồn tại và tầm quan trọng của tổ hợp từ, người học sẽ chủ động học khi bắt gặp chúng trong khi tiếp nhận ngôn ngữ đích.

Trong một đánh giá về các nghiên cứu thực nghiệm được thực hiện từ năm 2004 về tổ hợp từ tiếng Anh, Boers and Strong<sup>31</sup> đề xuất việc sắp xếp và tổ chức các hoạt động dạy tổ hợp từ trong sách giáo khoa nói chung theo 3 chiến lược: (1) thu hút sự chú ý của người học đến các tổ hợp từ được chọn làm trọng tâm, (2) khuyến khích người học tự tìm hiểu về tổ hợp từ qua từ điển hoặc các kho ngữ liệu, và (3) tăng cường sự tiếp xúc của người học với tổ hợp từ nhằm hỗ trợ cho việc ghi nhớ dài hạn. Mặc dù sự chú ý của người học là bước đầu quan trọng trong quá trình học từ vựng, phần lớn tổ hợp từ trong ngôn ngữ đầu vào lại không thu hút sự chú ý của người học vì tính rõ nghĩa của chúng<sup>32-34</sup> (ví dụ như *give advice* (cho lời khuyên) hay *do a test* (làm bài thi)). Một gợi ý được các học giả đề xuất nhằm giúp thu hút sự chú ý của người học đến các tổ hợp từ trong một bài đọc là bằng cách thay đổi định dạng trình bày của chúng (ví dụ như in đậm, gạch chân, in nghiêng, gạch nổi, hay thêm vào bản kẻ tổ hợp từ tương ứng bằng tiếng mẹ đẻ)<sup>17,35,36</sup>. Cách làm nổi bật bằng việc thay đổi định dạng trong văn bản được chứng minh là đem lại những tác động tích cực lên khả năng ghi nhớ của người học không chỉ đối với các từ vựng đơn lẻ<sup>35,37</sup>; mà còn đối với các tổ hợp từ<sup>38</sup>. Việc làm nổi bật các tổ hợp từ được chọn làm trọng tâm còn có tác dụng hướng người học đến việc tra cứu các cụm từ đó trong từ điển hay các kho ngữ liệu<sup>39</sup>. Mặc dù được giới thiệu như một giải pháp sơ phạm về việc xử lý tổ hợp từ trong sách giáo khoa, một số nghiên cứu thực nghiệm đã cho thấy rằng việc tra cứu tổ hợp từ không phải lúc nào cũng mang lại hiệu quả tích cực. Cụ thể là ở các nghiên cứu này hiệu quả học tập tổ hợp từ chỉ đạt ở mức dưới 40%<sup>40</sup> tới 60%<sup>41</sup>. Tuy nhiên, hiệu quả học tập thấp có thể là do sự tác động của một số yếu tố chẳng hạn như việc tra cứu không nhất quán, người học không được hướng dẫn cách tra cứu từ điển hoặc bị cản trở bởi sự quá tải về thông tin.

### **Việc tái sử dụng các tổ hợp từ đối với việc tiếp thu tổ hợp từ ở người học**

Rõ ràng là muốn thúc đẩy việc học từ vựng thì từ vựng phải được dạy và học một cách thích hợp<sup>27,42,43</sup>. Tuy nhiên, điều đó không có nghĩa là một khi đã được học thì người học sẽ nhớ và sẵn sàng cho việc sử dụng chúng bất cứ lúc nào. Nghiên cứu đã chỉ ra rằng việc tái sử dụng vốn từ vựng đã được dạy và học để nó lưu lại trong trí nhớ dài hạn của người học là tối quan trọng<sup>44</sup>. Thậm chí Nation<sup>45</sup> đã cảnh báo rằng việc dạy từ vựng mà không có biện pháp tái sử dụng thích hợp sẽ không mang lại những kết quả như mong đợi và như vậy mọi nỗ lực dạy sẽ trở nên vô ích. Do sự

khác biệt trong cơ hội cũng như môi trường thực hành từ vựng đã học, người học tiếng Anh như một ngoại ngữ cần tiếp xúc cùng một tổ hợp từ nhiều lần hơn so với người học tiếng Anh như một ngôn ngữ thứ hai<sup>32</sup>. Việc tái sử dụng để người học tiếp xúc một tổ hợp từ nhiều lần nên được thực hiện trong một khoảng thời gian kéo dài<sup>46,47</sup>. Tương tự, như đã được kiểm chứng trong một số nghiên cứu<sup>36,48-50</sup>, việc tiếp xúc nhiều lần một tổ hợp từ sẽ mang lại những tác động tích cực đối với việc tiếp thu từ vựng. Mặc dù các nghiên cứu này đưa ra các kết quả không đồng nhất về số lần cần tiếp xúc một tổ hợp từ để thúc đẩy tốt nhất việc học, chúng lại nhất quán ở một điểm là một tổ hợp từ càng được tiếp xúc thường xuyên thì càng dễ giúp người học tiếp thu và sẵn sàng cho việc sử dụng chúng. Các nghiên cứu này cũng chỉ ra rằng người học cần được tiếp xúc ít nhất 3 lần cùng một tổ hợp từ thì mới hy vọng là người học có thể nhớ và sử dụng chúng khi cần<sup>36,50</sup>.

### **PHƯƠNG PHÁP NGHIÊN CỨU**

Đây là một nghiên cứu dựa trên kho ngữ liệu quốc gia Anh ngữ (BNC). Để tài nghiên cứu này tập trung vào bộ sách tiếng Anh phục vụ cho việc dạy tiếng Anh theo chương trình của Bộ Giáo dục và Đào tạo do các nhóm tác giả trong đó có tác giả Hoàng Văn Văn làm chủ biên. Bộ sách có tên gọi là Tiếng Anh 1 đến Tiếng Anh 12. Ở mỗi cấp lớp, bộ sách được chia làm 2 quyển. Đối với cấp tiểu học, ở mỗi cấp lớp sách có 20 đơn vị bài học và sau mỗi 5 đơn vị bài học là bài ôn tập. Đối với cấp trung học cơ sở, ở mỗi cấp lớp sách có 12 đơn vị bài học và sau mỗi 2-3 đơn vị bài học là bài ôn tập. Đối với cấp trung học phổ thông, ở mỗi cấp lớp sách có 10 đơn vị bài học và sau mỗi 3 đơn vị bài học cũng có bài ôn tập. Bộ sách được thiết kế dựa trên cách tiếp cận tiếng Anh giao tiếp, tập trung chủ yếu vào việc phát triển kỹ năng nghe và nói cho học sinh. Bộ sách hiện đang trong giai đoạn sử dụng thử nghiệm ở nhiều trường công lập ở mọi cấp lớp.

Để trả lời cho các câu hỏi nghiên cứu, nhóm nghiên cứu thống kê các tổ hợp từ có cấu trúc động-danh và tính-danh trong bộ sách giáo khoa tiếng Anh để hiểu rõ các tổ hợp từ nào được sử dụng và được chọn là trọng tâm giảng dạy. Nhóm nghiên cứu bắt đầu bằng việc xây dựng kho ngữ liệu Anh ngữ từ bộ sách giáo khoa tiếng Anh nói trên. Sau đó, chúng tôi trích lọc tất cả các danh từ xuất hiện trong bộ sách. Tiếp đến, chúng tôi dùng chức năng gán nhãn POS-tagging trong BNC để trích xuất tất cả các tính từ và động từ nào kết hợp được với những danh từ vừa mới trích xuất. Sau khi trích xuất tất cả các cụm từ theo hai cấu trúc động-danh và tính-danh, chúng tôi kiểm tra chỉ số MI-score  $\geq 3$  và t-score  $\geq 4$  theo nghiên cứu của

Ackermann và Chen<sup>25</sup> để xác định cụm từ nào là tổ hợp từ.

Để trả lời câu hỏi nghiên cứu thứ nhất về mức độ tương thích giữa các tổ hợp từ được chọn làm trọng tâm trong bộ sách giáo khoa tiếng Anh và những danh mục tổ hợp từ được nêu trong phần lý thuyết tổng quan, đầu tiên chúng tôi nhận diện các tổ hợp từ mà bộ sách chọn làm trọng tâm giới thiệu đến người học. Theo trang Lời mở đầu của bộ sách, chúng là những tổ hợp từ được giới thiệu trong phần từ vựng của mỗi bài. Chúng cũng là những tổ hợp từ được lồng ghép trong phần nghe và đọc hiểu ở phần đầu của mỗi bài. Chúng tôi trích xuất tất cả các tổ hợp từ là trọng tâm của bộ sách ra một trang tính sau đó đối chiếu chúng với danh mục tổ hợp từ do Ackermann và Chen<sup>25</sup> xây dựng bởi vì như đã trình bày ở trên, đó là danh mục về tổ hợp từ học thuật có giá trị nhất. Tuy nhiên, nếu chỉ so sánh đối chiếu tổ hợp từ trọng tâm của bộ sách với danh mục tổ hợp từ học thuật của Ackermann và Chen<sup>25</sup>, sẽ khó có thể có cái nhìn toàn diện về hiệu quả của bộ sách trong việc giới thiệu các tổ hợp từ đến người học vì có rất nhiều tổ hợp từ tổng quát có giá trị trong việc dạy và học (ví dụ: *take a break, take action, fast food*) không xuất hiện trong danh mục do Ackermann và Chen<sup>25</sup> xây dựng. Như đã thảo luận ở phần tổng quan lý thuyết, vì không có danh mục tổ hợp từ tổng quát nào đáng để xem xét cho mục đích của nghiên cứu này, chúng tôi quyết định tiến hành thêm một bước. Đó là xét xem liệu các tổ hợp từ được chọn là trọng tâm trong bộ sách giáo khoa có phải là những tổ hợp từ có tần suất xuất hiện cao cần giới thiệu cho người học hay không. Thay vì xây dựng một danh mục tổ hợp từ tổng quát, chúng tôi đánh giá các tổ hợp từ được chọn là trọng tâm trong bộ sách thông qua một tiến trình gồm ba bước. Trước tiên, chúng tôi trích xuất các danh từ từ danh mục 2.801 từ danh mục từ vựng tổng quát (New General Service List - NGSL) vốn được Bộ giáo dục và Đào tạo mô tả là nền tảng cho việc thiết kế chương trình. Sau đó, chúng tôi kiểm tra xem các gốc danh từ của tổ hợp từ được chọn là trọng tâm trong bộ sách giáo khoa có xuất hiện trong danh mục NGSL hay không. Nếu không, các tổ hợp từ chứa những danh từ đó sẽ không được xem là tổ hợp từ có tần suất xuất hiện cao và do đó sẽ được kết luận là không đáng để tập trung vào giới thiệu cho người học. Nếu có, chúng sẽ được xử lý tiếp ở bước cuối cùng là kiểm tra tần suất xuất hiện ở BNC. Nếu chúng xuất hiện tối thiểu 30 lần trong kho ngữ liệu BNC, chúng là những tổ hợp từ có tần suất xuất hiện cao và vì vậy đáng được đưa vào chương trình giảng dạy.

Câu hỏi nghiên cứu thứ hai liên quan đến việc liệu các tổ hợp từ được chọn là trọng tâm trong bộ sách

giáo khoa có được tái sử dụng một cách có hệ thống và theo một nguyên tắc cụ thể để thúc đẩy việc học hay không. Để trả lời cho câu hỏi này, chúng tôi dò tìm và thống kê thủ công số lần xuất hiện của mỗi tổ hợp từ được chọn là trọng tâm trong bộ sách.

## KẾT QUẢ NGHIÊN CỨU

*Sự tương quan giữa tổ hợp từ trọng tâm trong bộ sách với danh mục tổ hợp từ tần suất cao*

Để trả lời câu hỏi này, trước hết chúng tôi xây dựng một kho ngữ liệu gồm 312.770 từ từ bộ sách giáo khoa tiếng Anh. Có 26.589 cụm từ có cấu trúc động-danh và 11.700 cụm từ có cấu trúc tính-danh được trích xuất. Sau đó, chúng được đối chiếu với BNC để kiểm tra chỉ số MI-score  $\geq 3$ , t-score  $\geq 4$ . Có 13.292 cụm từ động-danh và 11.079 cụm từ tính-danh đáp ứng các chỉ số và được nhận diện là tổ hợp từ. Bảng 1 dưới đây trình bày số lượng các dạng thức tổ hợp từ và hạng mục tổ hợp từ xuất hiện trong sách giáo khoa của từng cấp học. Các hạng mục tổ hợp từ và dạng thức tổ hợp từ được phân bố tăng dần theo từng cấp lớp, thể hiện sự nâng cao về số lượng từ vựng từ trình độ vỡ lòng lên trình độ trên trung cấp.

Theo kết quả đã trình bày ở trên, số lượng tổ hợp từ ở bộ sách dành cho cấp trung học cơ sở và trung học phổ thông hơn gấp nhiều lần số lượng tổ hợp từ dành cho cấp tiểu học. Điều này có thể do bộ sách dành cho cấp tiểu học tập trung chủ yếu vào các từ đơn lẻ và những câu đơn rất ngắn trong khi sách dành cho cấp trung học cơ sở cung cấp những bài đọc hiểu có nội dung dài hơn, khó hơn và đa dạng hơn về cả từ vựng lẫn cấu trúc. Vì thế, mức độ phân bố tổ hợp từ ở sách dành cho cấp tiểu học cũng thấp hơn so với mức độ phân bố tổ hợp từ ở sách dành cho cấp trung học cơ sở. Cụ thể là mật độ phân bố và tính đa dạng của tổ hợp từ tính trên tổng số 1.000 từ lần lượt là 19,35 và 10,88 (cao hơn mật độ phân bố và tính đa dạng của tổ hợp từ trong BNC mà Tsai (2005) đã thống kê<sup>17</sup>). Điều này cho thấy các tác giả của bộ sách giáo khoa đã có chủ ý đưa nhiều tổ hợp từ của ngôn ngữ đích vào phần từ vựng và bài tập của mỗi đơn vị bài học để người học tiếp xúc tổ hợp từ nhiều hơn.

Để xác định sự tương thích giữa các tổ hợp từ được chọn là trọng tâm với danh mục tổ hợp từ có giá trị cần được giới thiệu cho người học ngoại ngữ, chúng tôi trích lọc các tổ hợp từ mà bộ sách giáo khoa chọn làm trọng tâm. Như đã mô tả ở phần phương pháp, chúng là những tổ hợp từ được giới thiệu trong phần từ vựng và thường được làm nổi bật bằng cách gạch chân, tô đậm, in nghiêng trong bài đọc và nghe của mỗi đơn vị bài học. Có 542 tổ hợp từ động-danh và 536 tổ hợp từ tính-danh được nhận diện là những tổ hợp từ trọng tâm trong toàn bộ sách giáo khoa. Sau đó, chúng được

**Bảng 1: Phân bố hạng mục tổ hợp từ và dạng thức tổ hợp từ ở cấp học**

Sách	Số lượng hạng mục từ vựng	Số lượng hạng mục tổ hợp từ		Số lượng dạng thức tổ hợp từ	
		Đ-D	T-D	Đ-D	T-D
Cấp tiểu học	28.800	1.062	438	461	199
Cấp trung học cơ sở	142.580	6.109	4.828	4.024	3.101
Cấp trung học phổ thông	141.390	6.121	5.813	4.226	3.394
Tổng cộng	312.770	13.292	11.079	8.711	6.694

đối chiếu với danh mục tổ hợp từ học thuật do Ackermann và Chen<sup>25</sup> xây dựng gồm 1.773 tổ hợp từ tính-danh và 310 tổ hợp từ động-danh. Hình 1 bên dưới thể hiện số lượng các tổ hợp từ được chọn là trọng tâm và tỉ lệ phần trăm xuất hiện trong danh mục tổ hợp từ học thuật (ACL).

Như được trình bày ở Hình 1, có 45 trong tổng số 542 tổ hợp từ động-danh được chọn là trọng tâm và 174 trong số 536 tổ hợp từ tính-danh xuất hiện trong danh mục tổ hợp từ học thuật (ACL) có tần suất xuất hiện cao. Tổng cộng có một phần năm các tổ hợp từ được chọn là trọng tâm của bộ sách có xuất hiện trong ACL; tuy nhiên, xét về độ bao phủ thì số lượng tổ hợp từ được chọn là trọng tâm trong bộ sách giáo khoa chỉ chiếm 10,5% các tổ hợp từ của ACL. Điều này có nghĩa là bộ sách giáo khoa chưa trang bị tốt cho người học về tổ hợp từ học thuật.

Để có cái nhìn toàn diện về việc sách giáo khoa đã trang bị cho người học các tổ hợp từ tổng quát như thế nào, chúng tôi tìm hiểu liệu các tổ hợp từ được chọn là trọng tâm của bộ sách có phải là các tổ hợp từ có tần suất xuất hiện cao hay không. Điều này là dựa trên cơ sở lập luận rằng các tổ hợp từ có tần suất xuất hiện cao là các tổ hợp từ có giá trị và do đó đáng được giới thiệu đến người học một cách hợp lý<sup>31</sup>. Thật đáng ngạc nhiên khi thấy rằng một số lượng lớn tổ hợp từ (gồm 68 động-danh, 55 tính-danh) được giới thiệu như từ vựng mới một cách trùng lặp trong các bài khác nhau ở các quyển sách giáo khoa thuộc các cấp lớp khác nhau. Ví dụ như là tổ hợp từ *ride a bike*. Tổ hợp từ này được giới thiệu lần đầu tiên trong sách Tiếng Anh 2, bài 3 và sau đó lại được giới thiệu như là từ mới trong cuốn Tiếng Anh 4, bài 7 và cuốn Tiếng Anh 6, bài 12. Đối chiếu các gốc danh từ trong tổ hợp từ trọng tâm với danh mục các danh từ có tần suất cao trong NGSL, chúng tôi thấy 36 danh từ của tổ hợp từ động-danh và 26 danh từ của tổ hợp từ tính-danh được chọn trong bộ sách tiếng Anh không xuất hiện trong NGSL (xem phụ lục A). Nếu những gốc danh từ này không xuất hiện trong danh mục từ có tần suất xuất hiện cao thì các tổ hợp từ chứa chúng có thể xem là không xuất

hiện thường xuyên, do đó không nên ưu tiên dạy cho người học. Số lượng tổ hợp từ trọng tâm còn lại được đối chiếu với BNC để biết số lần xuất hiện ( $\geq 30$  lần). Hình 2 thể hiện số lượng tổ hợp từ không xuất hiện thường xuyên và những tổ hợp từ bị giới thiệu trùng lặp ở những bài khác nhau.

Có 123 trong tổng số 1.078 tổ hợp từ trọng tâm (11,4%) bị giới thiệu trùng lặp ở các đơn vị bài học nhau và 208 tổ hợp từ (19,3%) được xác định là không xuất hiện trong cả hai danh mục ACL và NGSL. Điều này có nghĩa là bộ sách giáo khoa chứa một số lượng lớn các tổ hợp từ không cần thiết giới thiệu đến người học.

#### *Việc tái sử dụng các tổ hợp từ trọng tâm*

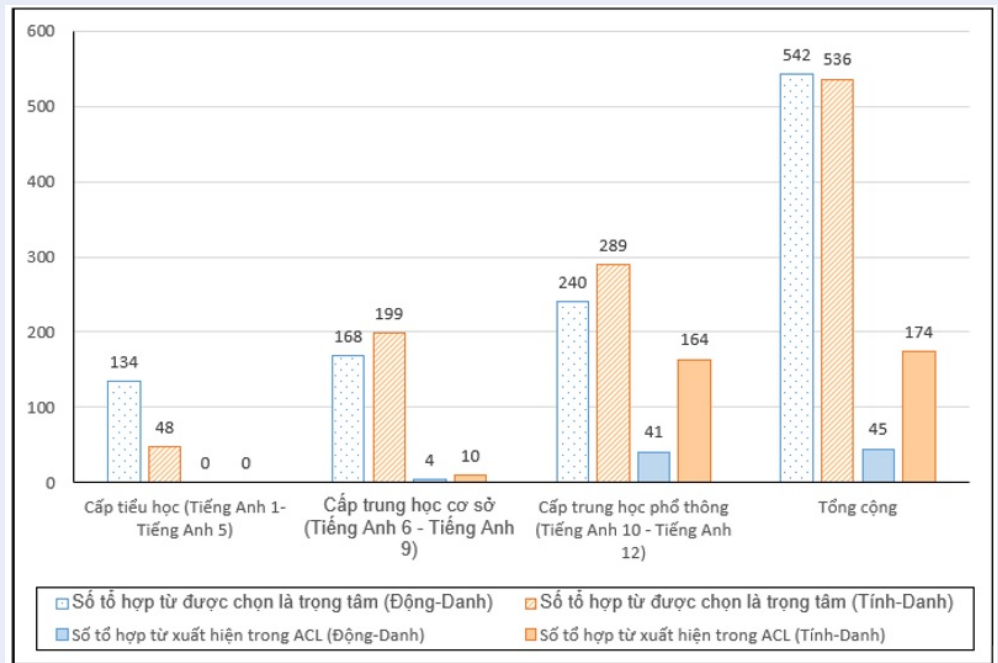
Để trả lời câu hỏi nghiên cứu cuối cùng, chúng tôi kiểm đếm số lượng tổ hợp từ trọng tâm tái xuất hiện theo từng nhóm. Hình 3 tóm tắt số lượng tổ hợp từ trọng tâm được tái sử dụng xuyên suốt cả bộ sách giáo khoa.

Như được trình bày ở Hình 3, số lượng tổ hợp từ trọng tâm có cấu trúc động-danh và tính-danh không hề được tái sử dụng là 658 tổ hợp từ (61%). 159 tổ hợp từ được tái sử dụng chỉ trong một đến hai bài. Và như đã nêu trong phần tổng quan lý thuyết, số ít lần gặp lại các tổ hợp từ như thế này không thật sự có ý nghĩa trong việc củng cố kiến thức của người học. Nếu ba lần là số lần tối thiểu mà một tổ hợp từ cần tái xuất hiện để giúp người học nhớ và sẵn sàng sử dụng chúng trong giao tiếp thì tổng số tổ hợp từ đáp ứng được mức tần suất này là 261 tổ hợp từ (24%). Tuy nhiên, cũng có một số tổ hợp từ tái xuất hiện trên mười lần xuyên suốt cả bộ sách giáo khoa chẳng hạn như *pay attention* (xuất hiện trong 20 bài), *do homework* (20 bài), *take turns* (14 bài), *social media* (11 bài), và *future generations* (11 bài).

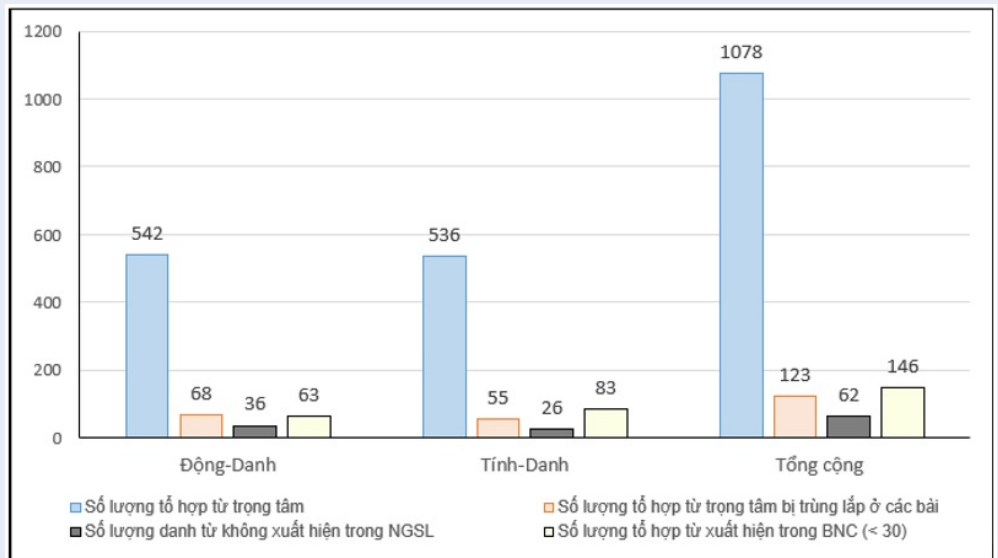
## **BÀN LUẬN**

### *Sự tương thích giữa tổ hợp từ trọng tâm trong bộ sách với các danh mục tổ hợp từ tham chiếu*

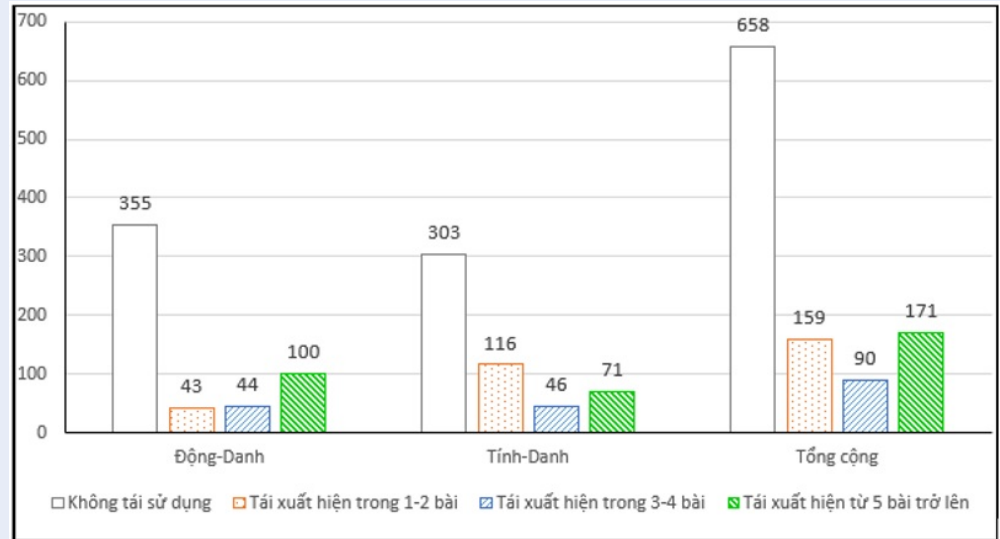
Theo kết quả đã trình bày ở trên, bộ sách có mức độ bao phủ các tổ hợp từ trong Danh mục tổ hợp từ học



**Hình 1:** Số lượng các tổ hợp từ được chọn là trọng tâm xuất hiện trong ACL (Lưu ý: ACL: Danh mục Tổ hợp từ Học thuật)



**Hình 2:** Số các tổ hợp từ trùng lặp và không xuất hiện thường xuyên



Hình 3: Số lượng tổ hợp từ trọng tâm được tái sử dụng trong các bài khác nhau

thuật (ACL) thấp. Điều này cho thấy người học chưa được trang bị đủ những tổ hợp từ học thuật để có thể sử dụng tiếng Anh cho mục đích học thuật ở cấp học cao hơn. Vì thế, nếu như học sinh ở cấp tiểu học và trung học cơ sở cần được trang bị nhiều các tổ hợp từ tổng quát thì học sinh ở cấp trung học phổ thông cần được trang bị nhiều các tổ hợp từ học thuật do ở cấp học này, người học đã quen dần với việc đọc những bài viết có nội dung chuyên sâu hơn và việc viết những bài văn thuộc nhiều thể loại khác nhau.

Nghiên cứu cũng cho thấy rằng dù có chứa khá nhiều tổ hợp từ có tần suất xuất hiện cao, bộ sách giáo khoa vẫn còn một số hạn chế nhất định như: nó chứa một số lượng đáng kể (31%) những tổ hợp từ trọng tâm được giới thiệu trùng lặp trong các bài khác nhau hoặc những tổ hợp từ không có tần suất xuất hiện cao hữu ích cho người học. Điều này chứng tỏ các tác giả của bộ sách giáo khoa cần đưa ra những tiêu chí thống nhất để lựa chọn và phân bố hợp lý hơn các tổ hợp từ trọng tâm ở các bài khác nhau trong từng quyển sách và xuyên suốt cả bộ sách. Như đã đề cập trong phần Tổng quan lý thuyết, việc tái sử dụng các tổ hợp từ đã giới thiệu sẽ mang lại những lợi ích quan trọng, giúp người học củng cố kiến thức về tổ hợp từ và sử dụng chúng một cách tự nhiên linh hoạt gần giống người bản ngữ<sup>30,31</sup>. Tuy nhiên, với thời lượng giờ học hữu ích ít ỏi trên lớp, việc dạy trùng lặp một số tổ hợp từ ở các bài khác nhau thực sự là một sự lãng phí. Một tổ hợp từ đã dạy trong một bài nên được tái sử dụng bằng cách lồng ghép vào những bài tập hay hoạt động khác nhau trong các bài tiếp theo xuyên suốt cả bộ

sách để việc củng cố kiến thức tổ hợp từ không gây nhàm chán và trùng lặp. Vì vậy, nhóm tác giả có thể cân nhắc biên soạn sách giáo khoa sao cho thời gian hữu ích trên lớp nên dành để giới thiệu thêm các tổ hợp từ cần thiết cho người học, tránh giới thiệu trùng lặp cùng một tổ hợp từ nhiều lần. Một giải pháp khác cho vấn đề này là nhóm tác giả nên có một lộ trình rõ ràng để kiểm soát tổ hợp từ trọng tâm nào được dạy ở bài nào. Việc soạn thêm một phụ lục danh mục tổ hợp từ cho toàn bộ sách thay vì chỉ giới thiệu tổ hợp từ trong danh mục từ vựng ở mỗi đơn vị bài học sẽ giúp người biên soạn sách và người sử dụng sách (bao gồm người học và người dạy) có cái nhìn tổng quát và toàn diện hơn về số lượng, phân bố và tần suất của những tổ hợp từ khi cần kiểm soát (người soạn), ôn tập (người học), hoặc tra cứu (người dạy). Có như vậy, nhóm tác giả bộ sách mới có thể vừa tránh được việc giới thiệu trùng lặp các tổ hợp từ trọng tâm ở các bài khác nhau, vừa kiểm soát được việc đưa các tổ hợp từ có tần suất xuất hiện cao vào mỗi bài theo một số lượng nhất định.

*Việc chọn lựa các tổ hợp từ làm trọng tâm để giới thiệu cho người học*

Bàn về số lượng các gốc danh từ tạo nên những tổ hợp từ trọng tâm của bộ sách, nghiên cứu cho thấy có 62 gốc danh từ không xuất hiện trong danh mục từ phổ thông có tần suất xuất hiện cao (NGSL). Bộ sách giáo khoa cũng chưa đặt trọng tâm vào các tổ hợp từ có gốc danh từ xuất hiện ở danh mục từ phổ thông NGSL (146 danh từ). Điều này có nghĩa là việc chọn lựa các tổ hợp từ nào làm trọng tâm để giới thiệu cho người



học chưa được ưu tiên và việc nên đưa tổ hợp từ nào vào sách nào ở cấp học nào chưa được xem xét trên những tiêu chí cụ thể rõ ràng. Mức độ cần thiết phải biết và sử dụng được tổ hợp từ, tần suất xuất hiện và giá trị học thuật của chúng cần được trình bày và phân biệt một cách đầy đủ và khoa học trong những hội đồng thẩm định sách giáo khoa các cấp. Nhìn chung, đây là vấn đề thường mắc phải của khâu biên soạn tài liệu học tập ngoại ngữ. Để xây dựng một bộ sách tốt về phương diện từ vựng, nhóm tác giả có thể xem xét thiết kế để cương từ vựng trước. Cụ thể là nên có các bước chọn lọc các tổ hợp từ hữu dụng cần giới thiệu đến người học giống như cách chọn lọc các từ vựng đơn lẻ. Ví dụ như nếu lượng từ vựng kỳ vọng đạt được ở một cấp học cụ thể là 1.000 từ có tần suất cao nhất thì tổ hợp từ trọng tâm sẽ phải chứa từ gốc là một trong những từ trong danh mục 1.000 từ này. Điều này là vì trong một số trường hợp, có những tổ hợp từ được xem là rất điển hình vì có chỉ số MI cao nhưng từ gốc trong tổ hợp từ lại nằm ngoài danh mục từ vựng cho người học ở cấp độ đó. Việc chọn các tổ hợp từ cũng phải xét đến trình độ từ vựng ngôn ngữ đích của người học và cấp độ mà người học cần đạt đến theo khung tiêu chuẩn được thiết kế cho từng bậc học dựa trên tham chiếu với khung trình độ chung Châu Âu (CEFR) hoặc khung năng lực ngoại ngữ 6 bậc của Việt Nam và quy đổi tương đương.

Mặc dù có những bằng chứng xác thực về tầm quan trọng của việc tái sử dụng từ vựng<sup>35-38</sup>, việc phân bố tần suất của các tổ hợp từ trong bộ sách giáo khoa dường như không đáp ứng được yêu cầu tái sử dụng hợp lý các tổ hợp từ. Chưa đến một phần tư các tổ hợp từ trọng tâm được tái sử dụng ba lần hoặc nhiều hơn xuyên suốt cả bộ sách, chứng tỏ là các tác giả của bộ sách giáo khoa chưa có một sự lưu ý thích đáng đến việc tái sử dụng tổ hợp từ. Phải thừa nhận rằng sẽ rất khó khi kỳ vọng các tác giả bộ sách giáo khoa cung cấp nhiều các tổ hợp từ đồng thời tái sử dụng chúng được nhiều lần như đã nêu trong phần Tổng quan lý thuyết. Hơn nữa, xuất phát từ lý thuyết rằng các tổ hợp từ tiếng Anh tương ứng với ngôn ngữ nguồn (ví dụ: *answer questions* = trả lời câu hỏi, *waste time* = lãng phí thời gian, *give information* = cung cấp thông tin) sẽ không gây ra khó khăn cho người học, vì vậy những tổ hợp từ tương ứng với ngôn ngữ nguồn không nhất thiết phải được tái sử dụng nhiều lần như với các tổ hợp từ không tương ứng. Thuận lợi của việc tác giả bộ sách giáo khoa có cùng tiếng mẹ đẻ với người học là họ có thể dễ dàng nhận diện tổ hợp từ nào không tương ứng với tiếng mẹ đẻ, và vì thế có thể tập trung vào việc tái sử dụng một cách thường xuyên hơn những tổ hợp từ này vốn được cho là hay gây ra khó khăn cho người học. Nhưng hạn chế của việc này chính là các tổ hợp

từ được đưa vào tái sử dụng có thể không có tần suất xuất hiện cao. Vì thế, việc trích lọc và tái sử dụng tổ hợp từ vẫn phải luôn xét đến mối tương quan giữa tần suất xuất hiện và tiêu chí MI.

Một vấn đề đáng lưu ý là phần lớn các tổ hợp từ trọng tâm chưa được tái sử dụng ở một tần suất nhất định đủ để giúp người học tiếp thu và sẵn sàng cho việc sử dụng chúng vì thời gian giảng dạy trên lớp đã bị lãng phí vào việc giới thiệu lại các tổ hợp từ đã được dạy. Và giải pháp cho vấn đề này là giáo viên có thể kết hợp thêm các tài liệu thích hợp khác để tăng thêm tần suất tiếp xúc với các tổ hợp từ cho học sinh, đặc biệt là những tài liệu mang tính chất dữ liệu tự nhiên được chính người bản xứ sử dụng trong giao tiếp (*authentic materials*). Một giải pháp khả thi khác đối với hạn chế này của bộ sách giáo khoa là nâng cao nhận thức của cả người dạy lẫn người học về tầm quan trọng của việc tái sử dụng các tổ hợp từ đã được dạy và học. Một khi đã nhận thức được tầm quan trọng của việc tiếp xúc thường xuyên các từ vựng nói chung và tổ hợp từ nói riêng thì giáo viên – những người tương tác trực tiếp với người học và sử dụng sách giáo khoa – có thể tái sử dụng chúng một cách sáng tạo thông qua các hoạt động tra đổi kỹ năng ngôn ngữ trên lớp, đặc biệt là hoạt động nói và viết. Việc đưa các tổ hợp từ vào để thi cũng được khuyến khích để có thể duy trì nhận thức về tầm quan trọng của tổ hợp từ trong việc dạy và học từ vựng. Làm được như thế, tầm quan trọng của tổ hợp từ sẽ được nhận thức một cách toàn diện và tác giả biên soạn sách giáo khoa cũng như giáo viên giảng dạy sẽ chú trọng đưa các tổ hợp từ vào mọi giáo trình, giáo án, để cương để nâng cao năng lực thẩm thấu và năng lực thể hiện ngôn ngữ cho người học. Ngoài ra, việc chọn lựa và giới thiệu các tổ hợp từ làm trọng tâm cũng cần được thực hiện một cách khoa học và nhất quán trong toàn bộ các sách thuộc bộ sách giáo khoa. Bên cạnh việc giới thiệu trực tiếp các tổ hợp từ qua danh mục từ vựng ở phần mở đầu bài học, qua các nội dung bài học, và qua bài tập sau mỗi bài học; các tác giả sách giáo khoa có thể xem xét bổ sung phần *Phụ lục Tổ hợp từ tổng hợp* để tiện cho người học tra cứu và ôn tập. *Phụ lục Tổ hợp từ Tổng hợp* này có thể giới thiệu các tổ hợp từ theo nhóm chủ đề thay vì theo trình tự xuất hiện như cách làm đối với *Phụ lục Tổ hợp từ* của từng sách riêng lẻ. Tác giả biên soạn sách giáo khoa có thể tham khảo các sách dạy từ vựng (phần tổ hợp từ) của các bộ sách English Collocations in Use để tham khảo các chủ đề cho từng cấp độ. Cần lưu ý là việc xây dựng *Phụ lục Tổ hợp từ tổng hợp* chỉ thích hợp cho phần tổng hợp tổ hợp từ ở cuối mỗi cấp học, không thích hợp cho từng sách riêng lẻ vì có thể không đủ số lượng tổ hợp từ để chia thành các nhóm tổ hợp từ theo chủ đề. Và việc phân chia này cũng chỉ

mang tính tương đối, sẽ có trường hợp một tổ hợp từ có thể được nhóm vào nhiều hơn một chủ đề hoặc không thuộc chủ đề nào cả. Trong trường hợp đó, tác giả biên soạn sách giáo khoa cần có chú giải để người sử dụng sách giáo khoa lưu ý.

Trong bối cảnh việc dạy ngoại ngữ ở Việt Nam đang có nhiều đổi mới, vai trò của giáo viên đứng lớp trong chương trình mới này càng quan trọng hơn bao giờ hết, nhất là khi Chương trình giáo dục phổ thông mới năm 2018 của Bộ Giáo dục và Đào tạo đặt trọng tâm vào chương trình thay vì sách giáo khoa. Như trên đã đề cập, giáo viên có thể quyết định giảng dạy, bổ sung hay lược bỏ một số nội dung trong sách giáo khoa cho phù hợp với trình độ của người học và tình hình cụ thể của việc dạy và học. Tuy sách giáo khoa không còn được coi là trọng tâm trong chương trình mới, nhưng nếu có thể, các tác giả các bộ sách giáo khoa tiếng Anh cũng nên xem xét các nội dung mà chúng tôi kiến nghị trong nghiên cứu này để cải tiến việc biên soạn tổ hợp từ trong bộ sách giáo khoa.

Trong khuôn khổ đề tài nghiên cứu này, chúng tôi chỉ dừng lại ở việc xem xét các tổ hợp từ được giới thiệu trong bộ sách giáo khoa có tương thích với các danh mục tổ hợp từ tham chiếu hay không. Các nghiên cứu sau này có thể cân nhắc các yếu tố khác liên quan đến việc trình bày tổ hợp từ trong bộ sách giáo khoa chẳng hạn như việc lựa chọn và thiết kế các dạng bài tập về tổ hợp từ nhằm tối ưu hóa quá trình nhận diện và ghi nhớ tổ hợp từ cho người học.

## KẾT LUẬN

Nghiên cứu này cho thấy rằng các tác giả bộ sách giáo khoa có dành sự chú ý nhất định cho các tổ hợp từ trong đó họ chú trọng vào việc gia tăng các hạng mục tổ hợp từ và dạng thức tổ hợp từ nhằm cung cấp cho người học một số lượng các tổ hợp từ thích hợp để sử dụng tương ứng với năng lực ngôn ngữ của họ. Tuy nhiên, tầm quan trọng của danh mục tổ hợp từ học thuật nên được xét đến trong quá trình tuyển chọn tổ hợp từ nào làm trọng tâm. Vì vậy, nghiên cứu này đề xuất xem xét bổ sung thêm các tổ hợp từ học thuật từ Danh mục Tổ hợp từ Học thuật (ACL) và tái sử dụng nhiều hơn tổ hợp từ tổng quát có tần suất xuất hiện cao. Tổ hợp từ học thuật có thể được giới thiệu nhiều hơn ở cấp học trung học phổ thông, khi mà học sinh đã có nền tảng kiến thức nhất định về ngoại ngữ. Đây cũng là một bước nhằm trang bị cho người học kiến thức tiếng Anh học thuật sẵn sàng cho việc học ở bậc đại học. Ngoài ra, để bộ sách giáo khoa được hoàn thiện hơn về phương diện từ vựng, nhóm tác giả có thể cân nhắc xây dựng giáo trình sao cho thời gian hữu ích trên lớp nên dành cho việc giới thiệu nhiều hơn các tổ hợp từ hữu dụng cho người học, tránh giới

thiệu trùng lặp cùng một tổ hợp từ. Một giải pháp khác cho vấn đề này là nhóm tác giả bộ sách nên có một lộ trình rõ ràng để kiểm soát tổ hợp từ trọng tâm nào được dạy ở bài nào. Có như vậy, nhóm tác giả bộ sách mới có thể không những tránh được việc giới thiệu lặp lại các tổ hợp từ trọng tâm ở các bài khác nhau mà còn bảo đảm các tổ hợp từ có tần suất xuất hiện cao được đưa vào mỗi bài theo một số lượng nhất định.

Việc chọn các tổ hợp từ điển hình nên xét đến trình độ từ vựng ngôn ngữ đích của người học. Ví dụ như nếu lượng từ vựng kỳ vọng đạt được ở một cấp học cụ thể là 1.000 từ có tần suất cao nhất thì tổ hợp từ trọng tâm sẽ phải chứa từ gốc là một trong những từ trong danh mục 1.000 từ này. Điều này là vì trong một số trường hợp, có những tổ hợp từ được xem là rất điển hình vì có chỉ số MI cao nhưng từ gốc trong tổ hợp từ lại nằm ngoài danh mục từ vựng cho người học ở cấp độ đó. Với bộ sách giáo khoa hiện thời, giáo viên được kỳ vọng đóng vai trò chủ động hơn trong việc tạo ra những hoạt động cho phép tái sử dụng các tổ hợp từ được chọn là trọng tâm một cách thú vị và có hệ thống hơn. Sách giáo khoa quan trọng bởi vì chúng có thể tối đa hoá tiếp xúc của người học với những từ vựng hữu dụng trong quá trình học tập. Giáo viên cũng quan trọng không kém vì họ đóng vai trò quyết định trong việc sử dụng sách giáo khoa một cách hiệu quả. Vì thế, hướng dẫn sử dụng sách giáo khoa hiệu quả cũng cần được phổ biến rõ ràng cho giáo viên. Mặc dù nghiên cứu này chỉ tập trung đánh giá việc biên soạn giới thiệu tổ hợp từ trong một bộ sách giáo khoa ở Việt Nam, nhưng các khuyến nghị cũng như đề xuất của nhóm nghiên cứu cũng có thể hữu ích cho các tác giả biên soạn sách giáo khoa tiếng Anh nói chung.

## LỜI CẢM ƠN

Nghiên cứu được tài trợ bởi Đại học Khoa học Xã hội và Nhân văn, Đại học Quốc Gia Thành phố Hồ Chí Minh trong khuôn khổ đề tài mã số TC2021-02/HĐĐT-ĐN&QLKH.

## DANH MỤC TỪ VIẾT TẮT

NGSL: danh mục Từ vựng phổ thông  
BNC: Kho ngữ liệu quốc gia Anh ngữ  
MI-score: chỉ số “cùng xuất hiện”  
ACL: danh mục tổ hợp từ học thuật

## XUNG ĐỘT LỢI ÍCH

Nhóm tác giả không có bất kỳ xung đột lợi ích nào trong công bố bài báo.

## ĐÓNG GÓP CỦA TÁC GIẢ

Tác giả Cao Thị Phương Dung thu thập, phân tích, xử lý dữ liệu; viết phần tóm tắt, giới thiệu, tổng quan lý thuyết, kết quả nghiên cứu.

Tác giả Phó Phương Dung phân tích, xử lý dữ liệu; viết phần phương pháp nghiên cứu.

Tác giả Đặng Nguyễn Anh Chi thu thập dữ liệu; viết phần thảo luận và kết luận.

## PHỤ LỤC A

Bảng 2.

## TÀI LIỆU THAM KHẢO

- Cao D, Deignan A. Using an online dictionary to support L2 learners' writing. In: Wright C, Harvey L, Simpson J, editors. *Voices and practices in applied linguistics: Diversifying a discipline*. York: White Rose University Press; 2019. P. 233-250; Available from: <https://doi.org/10.22599/BAAL1.n>.
- Nesselhauf N. The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*. 2003; 24: 223-242; Available from: <https://doi.org/10.1093/applin/24.2.223>.
- Siyanova A, Schmitt N. L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*. 2008; 64(3): 429-458; Available from: <https://doi.org/10.3138/cmlr.64.3.429>.
- Browne C, Culligan B, Phillips J. The New General Service List: A core vocabulary for EFL students and teachers. *JALT The Language Teacher*. 2013; 34(7): 13-15;.
- Hummel KM. *Introducing second language acquisition: Perspectives and practices*. Chichester, UK: Wiley Blackwell; 2014;.
- Clear J. From Firth principles: Computational tools for the study of collocation. In: Baker M, Francis G, Tognini-Bonelli E, editors. *Text and technology: In honour of John Sinclair*. Amsterdam: John Benjamins Publishing Company; 1993. P. 271-292; Available from: <https://doi.org/10.1075/z.64.18cle>.
- Durrant P. *High frequency collocations and second language learning*. Nottingham: University of Nottingham; 2008;.
- Sinclair J. *Corpus, concordance, collocation*. Oxford: Oxford University Press; 1991;.
- Stubbs M. *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell; 2002;.
- Hunston S. *Corpora in applied linguistics*. Cambridge: Cambridge University Press; 2002; Available from: <https://doi.org/10.1017/CBO9781139524773>.
- Jones S, Sinclair J. English lexical collocations. *Cahiers de Lexicologie*. 1974; 24: 15-61;.
- Durrant P, Schmitt N. Adult learners' retention of collocations from exposure. *Second Language Research*. 2010; 26: 163-188; Available from: <https://doi.org/10.1177/0267658309349431>.
- Stubbs M. Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics*. 2002; 7: 215-244; Available from: <https://doi.org/10.1075/ijcl.7.2.04stu>.
- Baker P. *Using corpora in discourse analysis*. London: Continuum; 2006; Available from: <https://doi.org/10.5040/9781350933996>.
- Gablasova D, Brezina V, McEnery AM. Collocations in corpus-based language learning research: identifying, comparing and interpreting the evidence. *Language Learning*. 2016; 67: 155-179; Available from: <https://doi.org/10.1111/lang.12225>.
- Kilgarriff A, Kosem I. *Corpus tools for lexicographers*. In: Granger S, Paquot M, editors. *Electronic lexicography*. Oxford: Oxford University Press; 2012. P. 31-55; Available from: <https://doi.org/10.1093/acprof:oso/9780199654864.003.0003>.
- Tsai K. Profiling the collocation use in ELT textbooks and learner writing. *Language Teaching Research*. 2015; 19: 723-740; Available from: <https://doi.org/10.1177/1362168814559801>.
- Bahns J, Eldaw M. Should we teach EFL students collocations? *System*. 1993; 21(1): 101-114; Available from: [https://doi.org/10.1016/0346-251X\(93\)90010-E](https://doi.org/10.1016/0346-251X(93)90010-E).
- Cowie AP. *Phraseology*. In: Ronald EA, editor. *The encyclopedia of language and linguistics*. Oxford: Pergamon; 1994. P. 3168-3171;.
- Nesselhauf N. *Collocations in a Learner Corpus*. Amsterdam: John Benjamin; 2005; Available from: <https://doi.org/10.1075/scl.14>.
- Benson M. The collocational dictionary and the advanced learner. In: Makhan T. *Learners' dictionaries: State of the art*. Singapore: SEAMEO Regional Language Centre; 1989. P. 84-93;.
- Laufer B. The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography*. 2010; 24: 29-49; Available from: <https://doi.org/10.1093/ijl/eqq039>.
- Shin D, Nation P. Beyond single words: The most frequent collocations in spoken English. *ELT Journal*. 2008; 62(4): 339-348; Available from: <https://doi.org/10.1093/elt/ccm091>.
- Chon YV, Shin D. A corpus-driven analysis of spoken and written academic collocations. *Multimedia-Assisted Language Learning*. 2013; 16(3): 11-38; Available from: <https://doi.org/10.15702/mall.2013.16.3.11>.
- Ackermann K, Chen, Y. Developing the Academic Collocation List (ACL) - A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*. 2013; 12: 235-247; Available from: <https://doi.org/10.1016/j.jeap.2013.08.002>.
- Lei L, Liu D. The Academic English Collocation List. *International Journal of Corpus Linguistics*. 2018; 23(2): 216-243; Available from: <https://doi.org/10.1075/ijcl.16135.lei>.
- Nation P. The four strands. *International Journal of Innovation in Language Learning and Teaching*. 2007; 1(1): 2-13; Available from: <https://doi.org/10.2167/ijlt039.0>.
- González FB, Schmitt N. How much collocation knowledge do L2 learners have? *ITL-International Journal of Applied Linguistics*. 2015; 166: 94-126; Available from: <https://doi.org/10.1075/itl.166.1.03fer>.
- Hsu JY. The effects of collocation instruction on the reading comprehension and vocabulary learning of Taiwanese college English majors. *Asian EFL Journal*. 2010; 12(1): 47-87;.
- Webb S, Newton J, Chang A. Incidental learning of collocation. *A Journal of Research in Language Studies*. 2013; 63: 91-120; Available from: <https://doi.org/10.1111/j.1467-9922.2012.00729.x>.
- Boers F, Strong B. An evaluation of textbook exercises on collocations. In: Tomlinson B, editor, *SLA research and materials development for language learning*. Oxfordshire: Routledge; 2016. P. 155-168; Available from: <https://doi.org/10.4324/9781315749082-22>.
- Hill J. Revising priorities: from grammatical failure to collocational success. In: Lewis M, editor, *Teaching collocation: Further developments in the lexical approach*. Boston, MA: Thomson Heinle; 2000. P. 145-157;.
- Lewis M. *Teaching collocation*. In: Lewis M, editor, *Teaching collocation: Further developments in the lexical approach*. Boston, MA: Thomson Heinle; 2000. P. 35-48;.
- Woolard G. *Collocation - encouraging learner independence*. In: Lewis M, editor, *Teaching collocation: Further developments in the lexical approach*. Boston, MA: Thomson Heinle; 2000. P. 132-145;.
- Nation ISP. *Teaching and learning vocabulary*. In: Hinkel E, editor. *Handbook of research in second language teaching and learning*. Mahwah, N.J.: Lawrence Erlbaum; 2005. P. 581-595;.
- Peters E. The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*. 2016; 20: 113-138; Available from: <https://doi.org/10.1177/1362168814568131>.

**Bảng 2: Các tổ hợp từ có chứa gốc danh từ không xuất hiện trong NGSL**

	Tổ hợp từ động-danh		Tổ hợp từ tính-danh	
1	make	trek	famous	astronauts
2	suffer	tragedy	red	balloon
3	draw	a diagram	promotional	brochure
4	play	badminton	used	can
5	play	basketball	national	certificate
6	join	bloc	harmful	chemicals
7	make	bows	household	chores
8	write	brochure	imperial	citadel
9	make	bundle	hazardous	chemicals
10	get	burn	cultural	complex
11	absorb	carbon dioxide	traditional	craft
12	explore	cave	subsequent	dynasties
13	describe	chart	sustainable	ecotourism
14	play	chess	new	emperor
15	write	CV	archaeological	excavations
16	keep	diary	leading	exporter
17	reduce	emissions	agricultural	exporter
18	explore	grottos	chemical	fertilizer
19	play	hide-and-peek	traditional	handcrafts
20	get over	jet lag	cognitive	impairments
21	do	karate	visual	impairments
22	experience	microgravity	digital	projector
23	have	picnic	academic	qualifications
24	go on	picnic	urban	sprawl
25	have	posters	spotted	squirrel
26	make	puppet	skilled	strategist
27	have	puzzle		
28	use	rickshaw		
29	build	sandcastles		
30	apply for	scholarship		
31	ride	scooter		
32	build	self-esteem		
33	play	table tennis		
34	share	viewpoints		
35	buy	roll		
36	make	rickshaw		

37. Goudarzi Z, Moini, M. R. The effect of input enhancement of collocations in reading on collocation learning and retention of EFL learners. *International Education Studies*, 2012; 5(3): 247-258; Available from: <https://doi.org/10.5539/ies.v5n3p247>.
38. Fazlali B, Shahini A. The effect of input enhancement and consciousness-raising techniques on the acquisition of lexical and grammatical collocation of Iranian EFL learners. *TESL-EJ*. 2019; 24(2);
39. Gürsoy, E. The noticing of formulaic sequences by Turkish learners of English. [MA dissertation]. Çukurova University; 2008;
40. Laufer B, Waldman T. Verb-noun collocations in second language writing: a corpus analysis of learners' English. *Language Learning*, 2011; 61: 647-672; Available from: <https://doi.org/10.1111/j.1467-9922.2010.00621.x>.
41. Komuro, Y. Japanese learners' collocation dictionary retrieval performance. In: Barfield A, Gilstad H, editors. *Researching collocations in another language*. London: Palgrave Macmillan; 2009. P: 86-98; Available from: [https://doi.org/10.1057/9780230245327\\_7](https://doi.org/10.1057/9780230245327_7).
42. Ellis R. Learner beliefs and language learning. *Asian EFL Journal*. 2008; 10(4): 7-25;.
43. Laufer B. Lexical frequency profiles: From Monte Carlo to the real world: A response to Meara (2005). *Applied Linguistics*. 2005; 26(4): 582-588; Available from: <https://doi.org/10.1093/applin/ami029>.
44. Brown R, Waring R, Donkaewbua S. Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*. 2008; 20(2): 136-163;.
45. Kojic-Sabo I, Lightbown PM. Students' approaches to vocabulary learning and their relationship to success. *The Modern Language Journal*. 1999; 83(2): 176-192; Available from: <https://doi.org/10.1111/0026-7902.00014>.
46. Harwood N. "At the end of the day they don't tally": What researchers tell us about gambits and how materials writers teach gambits. [MA dissertation]. Lancaster, UK: Lancaster University; 2000;.
47. Lewis M. *The lexical approach: The state of ELT and a way forward*. Hove, England: Language Teaching Publications; 1993;.
48. Rott S. The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*. 1999; 21(4): 589-619; Available from: <https://doi.org/10.1017/S0272263199004039>.
49. Webb S, Kagimoto E. Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics*. 2010; 32: 259-276; Available from: <https://doi.org/10.1093/applin/amq051>.
50. Webb S, Kagimoto E. The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*. 2009; 43: 55-77; Available from: <https://doi.org/10.1002/j.1545-7249.2009.tb00227.x>.

# An investigation into collocational profile in English textbooks for Vietnamese students

Cao Thi Phuong Dung\*, Pho Phuong Dung, Dang Nguyen Anh Chi



Use your smartphone to scan this QR code and download this article

## ABSTRACT

Collocations play a significant role in language teaching and learning. In an environment where English is taught as a foreign language, English textbooks are the major source of exposure for learners. Therefore, this study investigates (1) the compatibility of collocations targeted to high frequency collocations suggested in the literature, and (1) how collocations targeted are reused throughout the textbook series. An English corpus of 312,770 word tokens was built from the textbook series from which 13,292 collocations of verb-noun and 11,079 collocations of adjective-noun patterns were identified. The study found that the frequencies of occurrences of collocation tokens an collocation types increase gradually from one grade level to the next. Collocations that the textbook series intentionally introduce to the learner account for only 10,5% of the high frequency collocations in an academic collocation list. 11,4 % of the targeted collocations were duplicated in different units, and 19,3% were identified not high frequency collocations. 76% of collocations targeted are reused not to the point that learners can retain and be ready to use. Based on the findings, the study discusses some implications for collocation teaching and material designing.

**Key words:** collocation, EFL textbooks, collocation lists, vocabulary acquisition

University of Social Sciences and Humanities, VNUHCM, Vietnam

## Correspondence

**Cao Thi Phuong Dung**, University of Social Sciences and Humanities, VNUHCM, Vietnam

Email: ctpdung@hcmussh.edu.vn

## History

- Received: 28-11-2021
- Accepted: 20-5-2022
- Published: 30-6-2022

DOI : 10.32508/stdjssh.v6i2.723



## Copyright

© VNUHCM Press. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.



**Cite this article :** Cao, T. P. D., Phó, P. D., & Đặng, N. A. C. **An investigation into collocational profile in English textbooks for Vietnamese students.** *Sci. Tech. Dev. J. - Soc. Sci. Hum.* 2022, 6(2):1573-1586.