



PHÂN TÍCH VIỆC SỬ DỤNG NGỮ KẾT HỢP ĐỘNG TỪ-DANH TỪ TIẾNG ANH CỦA SINH VIÊN TRƯỜNG ĐẠI HỌC NGOẠI NGỮ, ĐẠI HỌC HUẾ, VIỆT NAM

Đặng Thị Cẩm Tú*, Thái Tôn Phùng Diễm, Ngô Thị Liên Hương, Trần Thị Minh Thi

¹Trường Đại học Ngoại Ngữ, Đại học Huế, 57 Nguyễn Khoa Chiêm, Tp. Huế

Tác giả liên hệ: Đặng Thị Cẩm Tú < camtu0601@gmail.com >

(Ngày nhận bài: 23-6-2021; Ngày chấp nhận đăng: 11-11-2021)

Tóm tắt. Nghiên cứu này phân tích mức độ phù hợp (appropriateness) của các ngữ kết hợp động từ-danh từ (verb-noun collocations) trong bài luận của sinh viên ngành Tiếng Anh tại Trường Đại học Ngoại ngữ, Đại học Huế so với các bài luận của sinh viên bản ngữ và giải thích kết quả thu được. Đầu tiên, các ngữ kết hợp động từ-danh từ sẽ được trích từ các bài luận của sinh viên bản ngữ và sinh viên Việt Nam. Dựa vào tần suất (frequency) và chỉ số tương hỗ (mutual information) trích xuất từ Corpus of Contemporary American English (COCA), chúng tôi xác định mức độ phù hợp của các ngữ kết hợp động từ-danh từ dựa theo 2 tiêu chí tần suất (freq) ≥ 6 , và chỉ số tương hỗ (MI) ≥ 3 (Siyanova & Schmitt, 2008). Số liệu định lượng cho thấy sinh viên Việt Nam sử dụng ngữ kết hợp động từ-danh từ phù hợp có vượt trội hơn sinh viên bản ngữ. Tuy nhiên, dữ liệu định tính từ phỏng vấn đã giúp giải thích việc sinh viên ngành Tiếng Anh sử dụng được các ngữ kết hợp tốt như vậy là vì các em có tham khảo các bài viết liên quan của các tác giả khác trên Internet chứ các em còn hạn chế trong việc tự sử dụng ngữ kết hợp để diễn tả ý tưởng của riêng mình.

Từ khóa: ngữ kết hợp, động từ - danh từ, tần suất, bài viết

AN ANALYSIS OF THE USE OF VERB-NOUN COLLOCATIONS IN STUDENTS' ESSAYS AT HUE UNIVERSITY OF FOREIGN LANGUAGES, VIETNAM

Dang Thi Cam Tu*, Thai Ton Phung Diem, Ngo Thi Lien Huong, Tran Thi Minh Thi

University of Foreign Languages, Hue University, 57 Nguyen Khoa Chiem St., Hue, Vietnam

* Correspondence to Dang Thi Cam Tu < camtu0601@gmail.com >

(Received: Juni 23, 2021; Accepted: September 11, 2021)

Summary. This study focused on analyzing the use of verb-noun collocations in argumentative essays among English-majored students at the Department of English, Hue University of Foreign Languages, Vietnam. Specifically, the study compared the appropriateness of verb-noun collocations used in argumentative essays of Vietnamese students with that of native speakers based on the two criteria: Mutual information ≥ 3 and Frequency ≥ 6 . Surprisingly, the findings showed that Vietnamese students used a higher number of appropriate verb-noun collocations in their essays than native speakers. However, interviews further revealed that Vietnamese students preferred to search for online writing pieces of the same topic and imitate collocations and ideas rather than make their own choices for lexical phrases and ideas, which, to some extent, helped explain their good use of verb-noun collocations in their essays.

Keywords: verb-noun collocations, essays, frequency, mutual information

1. Đặt vấn đề

Ngữ kết hợp, được định nghĩa là "đơn vị nhiều từ", hoặc "ngữ có sẵn" [5], là một phần không thể thiếu trong kiến thức ngôn ngữ của người học [15]. Việc sử dụng chính xác các ngữ kết hợp động từ-danh từ (verb-noun collocations, ví dụ: *make a mistake, do a favour, face a problem, gain knowledge*, vv.) đóng vai trò quan trọng trong việc nâng cao năng lực Tiếng Anh của người học. Việc sử dụng thành thạo các ngữ kết hợp trong Tiếng Anh giúp người học nói và viết Tiếng Anh lưu loát hơn [3]. Nghiên cứu của Abdi & Ariffin (2020) đã chỉ ra rằng khi người học sử dụng ngữ kết hợp phù hợp, bài viết của người học có khả năng đạt điểm cao hơn [1]. Nghiên cứu của Xiaolin (1993) cho thấy kiến thức ngữ kết hợp là một yếu tố quan trọng giúp việc giao tiếp của người học ngoại ngữ ở dạng văn viết đạt hiệu quả hơn, và chất lượng của ngữ kết hợp dùng trong bài viết (bao gồm độ đa dạng và sự chính xác) cũng góp phần làm nên chất lượng bài viết của người học [20]. Tuy nhiên, việc sử dụng chính xác và thành thạo các ngữ kết hợp này trong giao tiếp (ở dạng nói và viết) là điều không dễ dàng. Đã có nhiều nghiên cứu trên thế giới chỉ ra rằng, thậm chí những người học ngôn ngữ ở trình độ nâng cao vẫn gặp nhiều khó khăn khi sử dụng loại ngữ này [2], [13], [15].

Trong bối cảnh hiện nay, theo yêu cầu của Bộ Giáo dục và Đào tạo, sinh viên chuyên ngành Tiếng Anh cũng cần đạt được năng lực C1 theo định dạng VSTEP bậc 3-5 theo khung tham chiếu trình độ ngoại ngữ chung châu Âu. Để đạt được mức năng lực này, trong kỹ năng viết, sinh viên phải sử dụng được dải từ rất rộng một cách chính xác, bao gồm cả những từ ít gặp, đồng thời kiểm soát tốt phong cách viết và các cụm từ cố định (theo tiêu chí chấm thi môn Viết, Tài liệu hướng dẫn áp dụng Định dạng đề thi đánh giá năng lực sử dụng tiếng Anh từ bậc 3 đến bậc 5 theo Khung năng lực ngoại ngữ 6 bậc dùng cho Việt Nam trong việc xây dựng đề thi và chấm thi [6]). Do đó, việc nâng cao khả năng sử dụng chính xác các ngữ kết hợp của sinh viên trong bài viết luận là rất quan trọng.

Nghiên cứu này được thực hiện nhằm đánh giá mức độ chính xác, phù hợp của sinh viên khi sử dụng các ngữ kết hợp động từ - danh từ trong quá trình viết bài văn nghị luận tiếng Anh ở Đại học Ngoại ngữ, Đại học Huế. Bên cạnh đó, nghiên cứu hướng đến giúp sinh viên chuyên ngành Tiếng Anh (Đại học Ngoại ngữ, Đại học Huế) nhận thức được phần nào thực trạng sử dụng ngữ kết hợp động từ - danh từ của mình trong môn viết nghị luận, phát hiện ra những lỗi sai thường gặp và những điểm chưa phù hợp khi sử dụng các ngữ kết hợp động từ - danh từ. Từ đó, người nghiên cứu đưa ra những đề xuất để cải thiện việc dạy và học ngữ kết hợp.

2. Cơ sở lý luận

Trên thế giới đã có nhiều nghiên cứu về việc sử dụng ngữ kết hợp của người học ngoại ngữ so với người bản ngữ. Theo Henriken (2013), một số nghiên cứu cho thấy sự khác biệt trong cách sử dụng ngữ kết hợp giữa sinh viên bản ngữ và sinh viên học ngôn ngữ thứ hai [11]. Laufer và Waldman (2011) đã thực hiện một nghiên cứu cắt ngang (cross-sectional study) về cách sử dụng ngữ kết hợp động từ-danh từ trong 759 bài văn nghị luận của sinh viên bản ngữ tiếng Do Thái [13]. Trong nghiên cứu này, 220 danh từ xuất hiện nhiều nhất được trích từ Kho bài luận tiếng Anh bản ngữ Louvain Corpus of Native English Essays (LOCNESS), sau đó được lập thành danh sách và các ngữ kết hợp động từ-danh từ được trích xuất. Tần suất sử dụng các ngữ kết hợp động từ-danh từ giữa sinh viên bản ngữ và sinh viên phi bản ngữ được dùng như là một tiêu chí để so sánh. Sự khác biệt về tính chính xác của việc sử dụng ngữ kết hợp giữa sinh viên ở các mức độ thành thạo khác nhau cũng đã được khảo sát. Kết quả chỉ ra rằng sinh viên ở cả ba mức độ thành thạo (cơ bản, trung cấp và nâng cao) sử dụng ít ngữ kết hợp hơn nhiều so với sinh viên bản ngữ. Ngoài ra, lỗi sử dụng ngữ kết hợp vẫn tồn tại trong bài luận của sinh viên ở cả ba mức độ thành thạo. Ngoài so sánh cắt ngang này, nếu các nhà nghiên cứu tiến hành nghiên cứu sự phát triển của ngữ kết hợp trong một khoảng thời gian dài thì sẽ có thể có được một bức tranh toàn diện hơn.

Theo quan điểm dựa trên tần suất, Siyanova & Schmitt (2008) trích xuất 810 ngữ kết hợp tính từ-danh từ từ 31 bài luận của sinh viên Nga học tiếng Anh và 806 ngữ kết hợp tính từ-danh từ từ các bài luận của sinh viên bản ngữ trong LOCNESS (Louvain Corpus of Native English Essays) [17]. Các nhà nghiên cứu khảo sát tần suất (frequency) và chỉ số tương hỗ (MI – mutual information) của các ngữ kết hợp dựa vào Kho ngữ liệu Quốc gia Anh (BNC – British National Corpus) và thấy rằng trong nghiên cứu của họ, sinh viên đại học Nga sử dụng ngữ kết hợp tính từ-danh từ phù hợp (appropriate collocations) với tỷ lệ gần bằng tỷ lệ của sinh viên bản ngữ (45% so với 48%). Khả năng nhận định về tần suất của các ngữ kết hợp tính từ-danh từ và tốc độ xử lý ngữ kết hợp của sinh viên phi bản ngữ thấp hơn sinh viên bản ngữ. Tuy nhiên vì chỉ có 31 sinh viên Nga tham gia khảo sát nên Siyanova & Schmitt (2008) không đảm bảo kết quả nghiên cứu này đúng trên diện rộng [17]. Ngay cả bài phân tích kho ngữ liệu này cũng không kiểm soát được tác động của các chủ đề bài văn nghị luận.

Nessehaulf (2003) nghiên cứu 32 bài luận của sinh viên Đức năm thứ ba và năm thứ tư chuyên ngành Tiếng Anh tại một trường đại học [15]. Nessehaulf (2003) chấm các lỗi trong tất cả thành phần của ngữ kết hợp, bao gồm không chỉ động từ và danh từ, mà còn cả giới từ và mạo từ [15]. Nghiên cứu cho thấy lựa chọn sai các động từ là loại lỗi sai thường gặp nhất, tiếp theo là lựa chọn sai danh từ và lựa chọn sai các yếu tố phi từ vựng (ví dụ: *giới từ* và *mạo từ*). Điều này có nghĩa rằng người học cần biết cách sử dụng tất cả yếu tố từ vựng nhỏ nhất, chứ không phải chỉ riêng thành phần động từ, danh từ. Trong khi Laufer & Waldman (2011) chỉ tập trung vào yếu tố từ vựng của ngữ kết hợp động từ-danh từ (ví dụ: *throw a party* hay *pay attention*) [13], Nessehaulf (2003) sử dụng hướng tiếp cận toàn diện đối với ngữ động từ-danh từ, bao gồm các yếu tố từ ngữ và phi từ vựng (ví dụ: *raise the question of*, *provide a solution to*) [15]. Nessehaulf (2003) lập luận rằng, học viên cần biết cách sử dụng đúng toàn bộ yếu tố của một kết hợp ngữ. Bằng không, họ không thể tạo ra các ngữ kết hợp có thể chấp nhận được [15].

Ở Việt Nam, có một số nghiên cứu đã được thực hiện tập trung vào việc sử dụng ngữ kết hợp của sinh viên Việt Nam ở cấp đại học. Trần Thị Châu Pha và Nguyễn Thị Hà (2013) đã nghiên cứu trên 59 sinh viên năm thứ 2 ở một trường đại học thuộc Đồng Bằng Sông Cửu Long, sử dụng bài kiểm tra ngữ kết hợp (collocation test) và bảng hỏi (questionnaire) để xác định các lỗi sai phổ biến của sinh viên [16]. Kết quả nghiên cứu chỉ ra rằng sinh viên gặp nhiều khó khăn nhất trong việc sử dụng các ngữ kết hợp trạng từ-tính từ (ví dụ: *deeply impressive*) và ngữ kết hợp danh từ-danh từ (ví dụ: *a bar of chocolate*). Ngoài ra, nghiên cứu cũng cho thấy năng lực ngữ nghĩa (semantic competence) và năng lực văn hóa (cultural competence) là hai yếu tố quan trọng ảnh hưởng đến việc sử dụng ngữ kết hợp của sinh viên ngoại ngữ. Một nghiên cứu khác đã được thực hiện bởi Nguyễn Anh Thùy (2010), nghiên cứu việc sử dụng ngữ kết hợp của sinh viên năm thứ 3 chuyên ngành tiếng Anh ở trường Đại học Cần Thơ [19]. Tác giả cũng sử dụng một bài kiểm tra ngữ kết hợp và bảng hỏi để thu thập số liệu. Nghiên cứu chỉ ra rằng, kiến thức về ngữ kết hợp của sinh viên còn nhiều hạn chế. Trong các lỗi được tìm thấy, lỗi về ngữ kết hợp động từ-trạng từ là phổ biến nhất. Nghiên cứu cũng cho thấy việc dạy học là một yếu tố quan trọng ảnh hưởng đến khả năng sử dụng ngữ kết hợp của sinh viên.

Câu hỏi nghiên cứu

Do những nghiên cứu trước đây ở Việt Nam chưa tập trung vào vấn đề sử dụng ngữ kết hợp động từ-danh từ ở kỹ năng viết nên trong nghiên cứu này, tác giả sử dụng khối ngữ liệu để nghiên cứu việc sử dụng ngữ kết hợp động từ - danh từ tiếng Anh trong bài văn nghị luận của sinh viên Trường Đại học Ngoại ngữ, Đại học Huế thông qua việc so sánh mức độ phù hợp (appropriateness) của các ngữ kết hợp động từ-danh từ được sinh viên Việt Nam và sinh viên bản ngữ sử dụng.

Câu hỏi nghiên cứu: Sinh viên năm thứ ba ngành Tiếng Anh tại Trường Đại học Ngoại ngữ, Đại học Huế đã sử dụng ngữ kết hợp động từ-danh từ phù hợp như thế nào trong các bài văn nghị luận của mình so với sinh viên bản ngữ?

3. Phương pháp nghiên cứu

3.1. Phương pháp nghiên cứu chung

Câu hỏi nghiên cứu nhằm tìm ra mức độ phù hợp của các ngữ kết hợp động từ - danh từ được sinh viên Việt Nam sử dụng trong các bài văn nghị luận của họ và so sánh với sinh viên bản ngữ. Để trả lời được câu hỏi nghiên cứu này, chúng tôi sử dụng kết hợp cả phương pháp định lượng và định tính, kết hợp với một phần phương pháp khối ngữ liệu cho ngữ liệu bài luận tiếng Anh của sinh viên Việt Nam và sinh viên Mỹ. Về phương pháp định lượng, chúng tôi sử dụng phương pháp tiếp cận dựa trên tần suất và chỉ số tương hỗ, cùng với quy trình nghiên cứu của Siyanova & Schmitt [17]. Về phương pháp định tính, chúng tôi tiến hành phỏng vấn 4 sinh viên Việt Nam có bài luận được đem ra phân tích.

3.2. Phương pháp thu thập ngữ liệu và số liệu

Nguồn ngữ liệu: Khảo sát về cách sử dụng ngữ kết hợp động từ - danh từ của sinh viên Việt Nam trong nghiên cứu này được thực hiện dựa trên ngữ liệu các bài văn nghị luận do sinh viên ngành Tiếng Anh viết tại Trường Đại học Ngoại ngữ, Đại học Huế và các bài văn nghị luận của sinh viên Mỹ viết lấy từ khối ngữ liệu LOCNESS (Louvain Corpus of Native English Essays) ([8], [9]).

Khối ngữ liệu gồm 10 bài văn nghị luận của 10 sinh viên Việt Nam (mỗi bài dài khoảng 1.000 từ) được chọn ngẫu nhiên từ các bài luận của sinh viên năm thứ ba tại Trường Đại học Ngoại ngữ, Đại học Huế. Các sinh viên này đang theo học ngành Tiếng Anh (gồm cả sinh viên chuyên ngành Sư phạm Tiếng Anh và Ngôn ngữ Anh) và tham gia học phần Viết 5 – Viết văn nghị luận (argumentative essays) trong học kỳ một năm học 2019-2020. Học phần viết 5 là học phần viết cấp độ cao nhất tại Trường Đại học Ngoại ngữ, Đại học Huế. Sinh viên được viết về một đề tài nghị luận mà họ cảm thấy hứng thú và không bị áp lực về thời gian. Tổng số từ trong 10 bài văn nghị luận được chọn là 9.137 từ. Sinh viên được sử dụng các công cụ tham khảo trong suốt quá trình viết luận. Về ngữ liệu bản ngữ, 7 bài văn nghị luận (9.251 từ) do sinh viên Mỹ viết, được lấy từ Khối bài luận tiếng Anh bản ngữ LOCNESS. LOCNESS là khối ngữ liệu thu thập từ các bài luận do các sinh viên tại các trường đại học Anh và Mỹ viết [9], với tổng số từ lên đến 324.304 từ. Đầu tiên, tất cả các ngữ kết hợp động từ-danh từ trong các bài luận của sinh viên Việt Nam và Mỹ (ví dụ: *face a problem*) được trích xuất thủ công. Nếu các ngữ kết hợp này bao gồm tính từ đứng trước danh từ, thì chỉ trích phần lõi động từ-danh từ. Ví dụ, nếu cụm "*get good marks*" xuất hiện trong bài luận, thì chỉ trích cụm "*get marks*". Có tổng cộng 325 ngữ kết hợp động từ-danh từ được trích xuất từ các bài luận của sinh viên Việt Nam, và 327 ngữ kết hợp động từ - danh từ được trích từ các bài luận của sinh viên Mỹ.

Theo Granger & Tyson (1996), để so sánh các đặc điểm của sinh viên bản ngữ và sinh viên phi bản ngữ, chúng ta cần hai khối ngữ liệu tương đương [10]. Trong nghiên cứu này, ngữ liệu từ bài luận của sinh viên Việt Nam và bài luận của sinh viên Mỹ được chọn tương đương về độ dài (khoảng 9.000 từ mỗi bên), về thể loại (văn nghị luận), về độ tuổi (20-22 tuổi), không giới hạn thời gian viết cũng như việc được sử dụng công cụ tham khảo trong suốt quá trình thực hiện.

Ngoài ra, nhóm nghiên cứu tiến hành phỏng vấn 4 sinh viên khoa Tiếng Anh đã có bài viết được phân tích. Bước phỏng vấn này giúp nhóm nghiên cứu hiểu được cách sử dụng ngữ kết hợp động từ-danh từ của sinh viên và từ đó lý giải phần nào cho kết quả định lượng liên quan đến tần suất và chỉ số tương hỗ. Do mục đích chính của nghiên cứu này là đánh giá khả năng sử dụng ngữ kết hợp của sinh viên Việt Nam, từ đó đưa ra những kiến nghị cho việc dạy và học ngữ kết hợp cho sinh viên Việt Nam nên việc phỏng vấn chỉ thực hiện đối với sinh viên Việt Nam mà không thực hiện đối với sinh viên Mỹ.

3.3. Phương pháp phân tích ngữ liệu và số liệu

Theo Siyanova & Schmitt (2008) [17], mức độ phù hợp (appropriateness) của các ngữ kết hợp động từ-danh từ được đánh giá dựa trên hai tiêu chuẩn: (1) tần suất và (2) chỉ số tương hỗ (MI). Về tần suất kết hợp từ, chúng tôi áp dụng phân chia dài tần số của Siyanova & Schmitt (2008) [17]: tần suất ≤ 5 , 6-20, 21-100, và >100 lần. Về thông tin tương hỗ (MI), nó là thước đo thống kê khách quan được sử dụng để so sánh xác suất của hai yếu tố từ cùng đi với nhau (hay còn gọi là xác suất kết hợp) với xác suất khi quan sát từng yếu tố từ riêng biệt [7]. Theo Hunston (2002), điểm MI từ 3 trở lên thường biểu thị ranh giới trong việc xác định các ngữ kết hợp [12]. Chúng tôi áp dụng theo Siyanova & Schmitt (2008) [17], sử dụng ngưỡng tần suất là 6 và ngưỡng chỉ số tương hỗ là 3 để xác định mức độ phù hợp của các ngữ kết hợp động từ-danh từ trong nghiên cứu này. Nói cách khác, nếu một ngữ kết hợp động từ-danh từ được trích xuất trong COCA với tần số ≥ 6 và MI ≥ 3 , thì nó được đánh giá là một ngữ kết hợp động từ-danh từ phù hợp.

Khối ngữ liệu Corpus of Contemporary American English (gọi tắt là COCA) (<http://corpus.byu.edu/coca/>) được sử dụng để xác định tần suất và chỉ số tương hỗ của tất cả các ngữ kết hợp động từ-danh từ được trích xuất từ bài luận của sinh viên Mỹ và sinh viên Việt Nam. Chúng tôi chọn COCA vì các bài luận của sinh viên bản ngữ được dùng trong nghiên cứu này là của sinh viên Mỹ. COCA là một kho ngữ liệu lớn và đa dạng của tiếng Anh Mỹ với hơn 385 triệu từ được chia đều giữa nhiều thể loại (giao tiếp, tiểu thuyết, tạp chí nổi tiếng, báo và tạp chí học thuật). Nó được mô tả như một kho ngữ liệu "tổng quát" đang tìm cách phác họa một bức tranh toàn cảnh về ngôn ngữ [4].

Đối với số liệu thu được từ phỏng vấn, sau khi tiến hành phỏng vấn 4 sinh viên Việt Nam, nhóm nghiên cứu đã chuyển biên dữ liệu phỏng vấn và thảo luận để đưa ra các điểm

chính thu được từ phỏng vấn. Sau khi đã thống nhất các điểm chính, nhóm nghiên cứu lọc ra các đoạn trích liên quan để phục vụ cho việc phân tích sâu hơn.

4. Kết quả và thảo luận

Trong phần này, chúng tôi sẽ trình bày kết quả phân tích định lượng liên quan đến tần suất và chỉ số tương hỗ của các ngữ kết hợp động từ-danh từ được trích xuất từ bài luận của sinh viên Việt Nam và bài luận của sinh viên Mỹ. Từ đó, chúng tôi so sánh mức độ phù hợp trong việc sử dụng ngữ kết hợp giữa sinh viên Việt Nam và sinh viên Mỹ trong khi viết luận. Ngoài ra, chúng tôi cũng trình bày một số điểm chính trong kết quả phỏng vấn để giải thích thêm cho kết quả thu được từ việc phân tích định lượng.

4.1. Phân tích bài luận của sinh viên Việt Nam

Bảng 1. Phân bố các ngữ kết hợp được sinh viên Việt Nam sử dụng dựa trên cơ sở tần suất COCA (n=325)

	Tổng số ngữ kết hợp động từ-danh từ	Tần suất <=5	6<= Tần suất <=20	21<= tần suất <=100	100 < Tần suất
Số lượng ngữ kết hợp	325	27	42	87	169
%	100%	8%	13%	27%	52%

Bảng 1 cho ta thấy có tổng cộng 325 ngữ kết hợp động từ-danh từ được sử dụng trong 10 bài luận dài 9.137 từ của sinh viên năm thứ 3 khoa Tiếng Anh. Trong số 325 ngữ kết hợp này, tỉ lệ ngữ kết hợp có tần suất lớn hơn 100 trong khối ngữ liệu COCA chiếm tỉ lệ cao nhất, trên 50%. Tỉ lệ số ngữ kết hợp có tần suất nằm trong khoảng từ 6 đến 100 chiếm 40%. Tỉ lệ ngữ kết hợp có tần suất nhỏ hơn 6 chiếm 8%. Những ngữ kết hợp có tần suất rất cao có thể được kể đến như: spend time (tần suất: 5608), take time (tần suất: 4328), make friends (tần suất: 2887), give advice (tần suất: 2590), enjoy life (tần suất: 1169), make decision(s) (tần suất: 7592), make progress (tần suất: 1930), have access (tần suất: 9187), play (a) part (tần suất: 1642), have fun (tần suất: 10301). Để xác định độ phù hợp của các ngữ kết hợp, chỉ số tương hỗ được phân tích và kết quả được thể hiện ở bảng 2.

Bảng 2. Ngữ kết hợp động từ-danh từ có tần suất COCA ≥ 6 và MI là ≥ 3 được sinh viên Việt Nam sử dụng

	MI ≥ 3 FR ≥ 6	Tổng số ngữ kết hợp động từ-danh từ
N	138	325
%	42%	100

Bảng 2 cho thấy hai tiêu chí của mức độ phù hợp đã được giải thích trong phương pháp nghiên cứu, ngưỡng tần suất ≥ 6 và chỉ số tương hỗ ≥ 3 được sử dụng để xác định mức độ phù hợp của ngữ kết hợp động từ - danh từ trong nghiên cứu này. Khi cả tần suất và chỉ số tương hỗ được xét đến, kết quả cho thấy rằng 42% ngữ kết hợp sinh viên Việt Nam dùng trong bài luận đạt ngưỡng tần suất ≥ 6 và chỉ số tương hỗ ≥ 3 . Mặc dù sinh viên Việt Nam sử dụng một số ngữ kết hợp động từ - danh từ với tần suất phổ biến khá cao (trên 90% ngữ kết hợp có tần suất ≥ 6), khi cả tần suất và chỉ số tương hỗ đều được tính đến, dưới một nửa số ngữ kết hợp (42%) đáp ứng cả hai tiêu chí.

4.2. Phân tích bài luận của sinh viên bản ngữ

Bảng 3. Phân bố các ngữ kết hợp được sinh viên Mỹ sử dụng dựa trên tần suất COCA (n=327)

	Tổng số ngữ kết hợp động từ-danh từ	Tần suất ≤ 5	6 \leq Tần suất ≤ 20	21 \leq tần suất ≤ 100	100 < Tần suất
Số lượng ngữ kết hợp	327	34	49	77	167
%	100%	10%	15%	24%	51%

Bảng 3 cho thấy từ khối ngữ liệu 7 bài luận của sinh viên bản ngữ, 327 ngữ kết hợp động từ - danh từ được trích xuất, trong đó có hơn 50% cụm từ có tần suất >100 , 24% cụm từ có tần suất nằm trong khoảng từ 21-100, và 15% cụm từ có tần suất nằm trong khoảng 6-20. Ví dụ, trong ngưỡng >100 , có một số ngữ kết hợp như: have (a) date (tần suất: 1336), cause problem(s) (tần suất: 280), use technique(s) (tần suất: 269), read articles (tần suất: 3388), have (a) right (tần suất: 23161). Đối với ngưỡng 21-100, có một số ngữ kết hợp như sau: discover (a) cure (tần suất: 31), perform (a) test (tần suất: 66), find medication (tần suất: 21), use (a) survey (tần suất: 85), voice views (tần suất: 45). Trong ngưỡng tần suất từ 6-20, có một số ngữ kết hợp như: share obstacles (tần suất: 6), give result(s) (tần suất: 18), enhance meaning (tần suất: 15), fit activities (tần suất: 20).

Bảng 4. Ngữ kết hợp động từ - danh từ có tần suất COCA ≥ 6 và MI ≥ 3 được sinh viên Mỹ sử dụng

	MI ≥ 3 FR ≥ 6	Tổng số ngữ kết hợp động từ-danh từ
N	117	327
%	38%	100%

Bảng 4 cho thấy chỉ có 38% tổng số ngữ kết hợp được trích xuất từ bài luận của sinh viên Mỹ thỏa mãn được cả hai tiêu chí tần suất ≥ 6 và chỉ số tương hỗ ≥ 3 . Kết quả từ bảng 1, 2, 3 và 4 cho thấy rằng số lượng ngữ kết hợp động từ - danh từ được sinh viên bản ngữ sử dụng trong bài viết của mình hầu như gần bằng với số lượng ngữ kết hợp động từ - danh từ được sinh viên khoa Tiếng Anh – Đại học Ngoại ngữ, Đại học Huế (327 so với 325) sử dụng. Số lượng ngữ kết hợp động từ - danh từ của sinh viên bản ngữ có tần suất xuất hiện trong COCA >100 chiếm khoảng 50%, tương đương với sinh viên Việt Nam. Tuy nhiên, đối với các ngữ kết hợp động từ - danh từ đáp ứng cả hai tiêu chí Freq ≥ 6 và MI ≥ 3 thì sinh viên bản ngữ lại dùng ít hơn trong bài luận của mình (38%) so với sinh viên Việt Nam (42%). Phát hiện này dường như mâu thuẫn với giả thuyết rằng người bản ngữ có khả năng sử dụng nhiều ngữ kết hợp tốt hơn người học tiếng Anh. Số liệu thu được này cũng đi ngược lại với kết quả của một số nghiên cứu trước đây, ví dụ như Laufer và Waldman (2011) [13] và Siyanova & Schmitt (2008) [17].

Mặc dù đây chỉ mới là một nghiên cứu có quy mô nhỏ, nhưng nó cũng cho chúng ta một góc nhìn về việc sử dụng ngữ kết hợp của sinh viên Việt Nam. Đầu tiên, có vẻ đáng ngạc nhiên khi sinh viên Việt Nam sử dụng ngữ kết hợp động từ - danh từ trong bài luận nhiều hơn sinh viên bản ngữ. Điều này có thể được giải thích bởi thực tế là các sinh viên năm thứ ba trong nghiên cứu này có trình độ ngôn ngữ khá cao; do đó họ có xu hướng tự tin về kiến thức ngôn ngữ của mình [13] và có xu hướng sử dụng càng nhiều ngôn ngữ càng tốt trong bài luận của họ. Thứ hai, nghiên cứu này cho thấy sinh viên năm thứ 3 trường Đại học Ngoại ngữ, Đại học Huế sử dụng nhiều ngữ kết hợp phù hợp (ngữ kết hợp thỏa mãn 2 điều kiện: Freq ≥ 6 và MI ≥ 3) hơn sinh viên bản ngữ. Vì kết quả này khá ngạc nhiên, nhóm nghiên cứu đã đi tìm nguyên nhân trong phần phỏng vấn các sinh viên Việt Nam và nhận được một số lý giải.

Trong số 4 sinh viên được phỏng vấn, cả 4 em đều nói rằng các em không hề dùng bất cứ công cụ gì để tra khảo ngữ kết hợp trong khi viết luận, mặc dù quá trình viết luận này được làm ở nhà và các em biết là được phép dùng tài liệu tham khảo. Các em cũng nói là không có thói quen tra khảo ngữ kết hợp trong khi viết luận.

“Em thường không tra khảo cụm collocations và cũng không ý thức được mình dùng đúng hay sai. Em cũng không hề nghĩ đến chuyện phải tra khảo ngữ kết hợp trong khi viết.” (sinh viên T)

“Thường thì em tự dùng ngữ kết hợp nào em biết, hoặc em nhớ ngữ nào thì em dùng lại ngữ đó thôi, chứ em không tra khảo các ngữ trong khi viết. Mặc dù viết ở nhà nhưng em viết rất nhanh, không cần nhắc nhiều.” (Sinh viên Y)

Sinh viên cũng giải thích thêm là khi viết bài về một chủ đề nào đó, thay vì tra ngữ kết hợp, các em thường tra ý tưởng ở Google hoặc Google Scholars. Sinh viên phản ánh trong phần phỏng vấn rằng các em có thói quen tra ý tưởng ở trên mạng Internet rồi viết lại, chứ ít khi tra từng ngữ kết hợp. Khi đọc các ý tưởng trên mạng, các em thấy một số ngữ người ta đã dùng để diễn tả ý và dùng lại các ngữ đó để diễn đạt ý chứ không dùng ngữ mới. Điều này phần nào giải thích được vì sao việc sinh viên Việt Nam sử dụng nhiều ngữ kết hợp động từ - danh từ trong bài luận của họ theo số liệu định lượng thu được ở bảng 1, 2, 3, 4. Nghiên cứu của Laufer (2011) đã chỉ ra rằng phần lớn người học cảm thấy không cần tra khảo ngữ kết hợp khi đang viết, vì điều đó ảnh hưởng đến mạch suy nghĩ trong khi viết [14]. Tarp và cộng sự (2017) cũng chỉ ra rằng trong quá trình viết luận, nếu người học sử dụng các nguồn tài liệu tham khảo để tra ngữ kết hợp, lúc quay lại viết, khả năng cao là họ mất tập trung và quên những suy nghĩ diễn ra trong đầu họ trước khi tra cứu [18].

Khi được hỏi có biết sử dụng khối ngữ liệu (corpus) để hỗ trợ cho việc sử dụng ngữ kết hợp trong khi viết luận ở nhà không, sinh viên đều nói chưa từng nghe khái niệm corpus. Các em cũng phản ánh là rất ít khi dùng từ điển để tra khảo ngữ kết hợp.

“Khi viết luận ở nhà, em search trên google các bài viết liên quan đến chủ đề mình sẽ viết, đọc qua và lấy ý tưởng, rồi dùng theo các ngữ mà họ dùng để diễn đạt ý đó thôi chứ không dùng ngữ mới.” (Sinh viên T)

“Mỗi lần viết bài luận hay thậm chí viết tiểu luận thì em chỉ tập trung viết trong một ngày nên ít tra từ điển lắm. Em cũng không có kinh nghiệm dùng các công cụ tra khảo ngữ kết hợp như corpus mà cô đã đề cập.” (Sinh viên B)

Từ kết quả này, chúng ta phần nào thấy được rằng mặc dù sinh viên Việt Nam sử dụng nhiều ngữ kết hợp phù hợp trong bài luận, nhưng nhận thức của các em về việc sử dụng các ngữ kết hợp đó trong khi viết luận còn rất hạn chế. Những thói quen và cách tiếp cận sao chép hay tham khảo ý tưởng sau đó sử dụng lại các ngữ kết hợp người khác đã sử dụng trong tài liệu của họ hiện đang khá phổ biến (4/4 sinh viên được phỏng vấn đều nêu ra thói quen như vậy). Điều này phản ánh phần nào thói quen và tư duy sao chép vẫn còn tồn tại trong sinh viên khoa Tiếng Anh hiện nay ở Trường Đại học Ngoại ngữ, Đại học Huế. Sinh viên vẫn còn chưa có ý thức nhiều về việc rèn luyện khả năng tự sử dụng ngôn ngữ để diễn đạt ý tưởng của riêng mình. Điều này cũng lý giải được phần nào tại sao sinh viên học tiếng Anh lại có thể đạt được tỉ lệ sử dụng ngữ kết hợp trội hơn sinh viên bản ngữ (kết quả số liệu định lượng thu được ở bảng 1, 2, 3, 4).

5. Kết luận và đề xuất

Nghiên cứu này đã nêu ra một thực trạng đang tồn tại trong sinh viên khoa Tiếng Anh – Trường Đại học Ngoại ngữ, Đại học Huế hiện nay, đó là khi viết luận Tiếng Anh, các em chưa thực sự độc lập trong suy nghĩ và trong việc sử dụng các ngữ kết hợp. Khi so sánh với sinh viên bản ngữ, sinh viên Việt Nam vẫn trội hơn về việc sử dụng các ngữ kết hợp động từ-danh từ trong bài văn nghị luận của họ, nhưng đó chủ yếu là nhờ vào việc các em tham khảo các bài viết của người khác ở trên Internet và bắt chước, chứ việc tự suy nghĩ các ngữ kết hợp để diễn đạt được ý tưởng của riêng các em là điều còn rất hạn chế. Các em cũng chưa tiếp cận nhiều với việc sử dụng các ngữ liệu trực tuyến như khối ngữ liệu (corpus) hay từ điển (online và bản giấy) để tham khảo các ngữ kết hợp và tự dùng các ngữ kết hợp cho chính xác. Tuy nhiên, vì nghiên cứu này mới chỉ dừng lại ở việc phân tích khoảng 9.000 từ lấy từ bài luận của sinh viên khoa Tiếng Anh và chỉ mới phỏng vấn 4 sinh viên nên kết quả nghiên cứu chưa có tính khái quát cao. Vì thế, đây là một hạn chế mà nhóm nghiên cứu mong muốn các nghiên cứu sau này tiếp tục phát triển để phân tích số lượng bài luận của sinh viên khoa Tiếng Anh – Đại học Ngoại ngữ, Đại học Huế ở diện rộng hơn. Mặt khác, nghiên cứu này cũng chỉ mới giới hạn ở khoa Tiếng Anh, Trường Đại học Ngoại ngữ, Đại học Huế nên nhóm nghiên cứu mong muốn nghiên cứu được mở rộng hơn trên quy mô toàn quốc để có một kết quả đáng tin cậy hơn. Từ kết quả nghiên cứu trên, nhóm nghiên cứu cũng có một số đề xuất liên quan đến việc rèn luyện cho sinh viên tư duy độc lập trong việc sử dụng các ngữ kết hợp khi viết, ví dụ, giáo viên có thể hướng dẫn và cho sinh viên trải nghiệm việc dùng khối ngữ liệu Contemporary Corpus of American English (COCA) hay khối ngữ liệu British National Corpus (BNC) để tra khảo độ phù hợp, độ chính xác của các ngữ kết hợp. Ngoài ra, nhóm nghiên cứu cũng thiết nghĩ, nên có các phần mềm hỗ trợ phát hiện lỗi đạo văn để từ đó khuyến khích, thúc đẩy việc sinh viên tự sáng tạo, tự suy nghĩ ý tưởng cho bài luận của mình cũng như nâng cao khả năng sử dụng ngữ kết hợp của các em trong khi viết luận tiếng Anh.

TÀI LIỆU THAM KHẢO

1. Abdi, B. M., & Ariffin, A. (2020). The relationship between collocation competence and writing skills of EFL learners. *The Asian Journal of English Language and Pedagogy*, 8(1), 41-52.
2. Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations?. *System*, 21(1), 101-114.
3. Bazzaz, F. E. (2016). Teaching verb-noun collocations is to kill two birds with one stone. *International Journal on Studies in English Language and Literature*. 4(10), 73-81.

4. Bennett, G. R. (2010). *Using Corpora in the Language Learning Classroom: Corpus linguistics for teachers*. Ann Arbor, MI: University of Michigan Press.
5. Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, 10(3), 245 – 261.
6. Bộ Giáo dục và Đào tạo, Tài liệu “*Hướng dẫn áp dụng Định dạng đề thi đánh giá năng lực sử dụng tiếng Anh từ bậc 3 đến bậc 5 theo Khung năng lực ngoại ngữ 6 bậc dùng cho Việt Nam trong việc xây dựng đề thi và chấm thi*” (Phê duyệt kèm theo Quyết định số: 730/QĐ-BGDĐT ngày 11 tháng 3 năm 2015 của Bộ trưởng Bộ Giáo dục và Đào tạo).
7. Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22 – 29.
8. Granger, S., Sanders, C. & Connor, U. (n.d.). *LOCNESS: Louvain corpus of native English essays*. Truy cập 15/11/2007 từ <http://www.fltr.ucl.ac.be/fltr/germ/etan/cecl/CeclProjects/Icle/locness1.htm>.
9. Granger, S. & DeCock, S. (2004). *Introduction to The Louvain Family of Corpora: ICLE, LOCNESS, LOCNEC, LINDSEI, LINDSEI – the German component*. Catholic University of Louvain: Belgium.
10. Granger, S. & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15(1), 17 – 27.
11. Henriksen, B. (2013). Research on L2 learners’ collocational competence and development a progress report. In C. Bardel, C. Lingquist & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use. New perspectives on assessment and corpus*, (pp.29–56). European association of second Language Acquisition Monographs. Uppsala University Publications.
12. Hunston (2002). *Corpora in applied linguistics*. Cambridge, UK: Cambridge University Press.
13. Laufer, B. & Waldman, T. (2011). Verb-noun collocations in second language writing: a corpus analysis of learners’ English. *Language learning*, 61(2), 647 – 672.
14. Laufer, B. (2011) The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography*, 24(1), 29–49.
15. Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223 – 242.

16. Trần Thị Châu Pha & Nguyễn Thị Hà (2013). Students' ability in using English collocations: An analysis of Vietnamese EFL learners. *Journal of Science and Technology – The University of Danang*, 2(63), 79-86.
17. Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review*, 64(3), 429 – 458.
18. Tarp, S., Fisker, K. Sepstrup, P. (2017) L2 writing assistants and context-aware dictionaries: New challenges to lexicography. *Lexikos*, 27, 494–521.
19. Nguyễn Anh Thùy (2010). *CTU English majored students' ability in using English collocations*. Cần Thơ: Việt Nam.
20. Xiaolin, Z. (1993). *English collocations and their effect on the writing of native and non-native college freshmen*. Doctoral Thesis. Indiana University of Pennsylvania.