

THIẾT KẾ BÀI TẬP LIÊN KẾT ĐỘNG TỪ-DANH TỪ TIẾNG ANH DỰA TRÊN KHỐI NGỮ LIỆU

Phan Thị Thanh Thảo

Trường Đại học Ngoại ngữ, Đại học Huế

pttthao@hufliis.edu.vn

(Nhận bài: 20/12/2023; Hoàn thành phản biện: 29/02/2024; Duyệt đăng: 25/04/2024)

Tóm tắt: Ứng dụng phần mềm quản lý kho ngữ liệu vào việc phát triển nguồn tài liệu giảng dạy ngôn ngữ đã trở thành một trong những lĩnh vực nghiên cứu tích cực trong Ngôn ngữ học máy tính (CL). Bài báo này trình bày việc xây dựng bài tập liên kết động từ - danh từ (V-N collocations) thường gặp trong tiếng Anh với sự hỗ trợ của máy tính dựa trên Khối ngữ liệu BNC và WordNet cho người học tiếng Anh ở Việt Nam. Với phần mềm Sketch Engine và khối ngữ liệu BNC và Wordnet, chúng tôi đã xây dựng khoảng 500 bài tập trắc nghiệm để giảng dạy tiếng Anh cho sinh viên ngành ngôn ngữ Anh năm thứ nhất. Từ những kết quả thu được, chúng tôi đề xuất về việc vận dụng khối ngữ liệu vào việc xây dựng tài liệu giảng dạy ngữ pháp nói chung và kết hợp từ trong tiếng Anh nói riêng nhằm nhấn mạnh tầm quan trọng của ứng dụng ngành Xử lý ngôn ngữ tự nhiên vào việc dạy học tiếng Anh cũng như các ngoại ngữ khác tại Việt Nam.

Từ khóa: Khối ngữ liệu, liên kết danh từ-động từ, bài tập trắc nghiệm

1. Giới thiệu

Đối với những người học tiếng Anh như một ngoại ngữ hay ngôn ngữ thứ 2 (L2), việc nắm vững và sử dụng đúng liên kết động từ - danh từ (V-N collocations) trong tiếng Anh thường đem lại nhiều thách thức do có sự khác biệt về cấu trúc và ngữ nghĩa, cũng như sự kết hợp đa dạng của từ vựng trong ngữ cảnh khác nhau giữa tiếng mẹ đẻ (L1) và ngôn ngữ thứ hai (L2). Chẳng hạn, một số liên kết động từ - danh từ có sự tương đồng về nghĩa từ vựng giữa tiếng Anh và tiếng Việt (do housework = làm việc nhà; pay a visit = viếng thăm), tuy nhiên, rất nhiều cụm từ liên kết có nghĩa hoàn toàn khác giữa hai ngôn ngữ này, ví dụ, *kick (đá) the buck* (cái xô) có nghĩa là *chết*, *qua đời* chứ không tương đồng với nghĩa đen “*đá cái xô*” trong tiếng Việt; *hit the hay* nghĩa là *đi ngủ* chứ không liên quan gì đến nghĩa *đập cỏ khô*. Vì vậy, việc sử dụng đúng các liên kết từ đóng vai trò quan trọng trong giao tiếp và truyền đạt thông tin từ một ngôn ngữ này sang một ngôn ngữ khác.

Nhằm giúp người học tiếng Anh sử dụng chính xác các liên kết từ, đặc biệt liên kết danh từ - động từ, một liên kết khá phổ biến trong ngôn ngữ này, các nhà ngôn ngữ và người giảng dạy ngôn ngữ luôn tìm cách để xây dựng nhiều tài liệu để người học có thể luyện tập thường xuyên nhằm tránh lỗi sai. Cùng với sự phát triển mạnh mẽ của ngành ngôn ngữ học máy tính, có thể thấy rằng việc xây dựng tài liệu giảng dạy và học tập dựa trên ứng dụng ngôn ngữ học khối ngữ liệu là một trong những giải pháp hiệu quả, tiết kiệm thời gian và công sức. Tuy nhiên, đối với những nhà nghiên cứu ngôn ngữ và các giáo viên tiếng Anh, việc xây dựng hệ thống bài tập về liên kết từ trong tiếng Anh đòi hỏi sự hiểu biết vận dụng cao kiến thức công nghệ thông tin và ngôn ngữ học máy tính vào thực tiễn giảng dạy ngôn ngữ tiếng Anh. Hơn nữa, theo tìm hiểu của chúng tôi, ngoài các nghiên cứu về ứng dụng ngôn ngữ học khối ngữ liệu trong nghiên cứu và giảng dạy ngoại ngữ qua ví dụ tiếng Đức (Đặng, 2021), và trong việc dạy học từ vựng tiếng Việt cho người nước ngoài qua ngữ liệu song ngữ Anh-Việt (Lê & Đỗ, 2021), vẫn chưa có nghiên cứu nào tại các khoa

ngoại ngữ khác về việc ứng dụng Ngôn ngữ học khối liệu vào xây dựng tài liệu dạy học như bài luyện tập cách sử dụng liên kết từ tiếng Anh cho người học.

Với những vấn đề được nêu ra ở trên, chúng tôi đề xuất việc xây dựng một số bài tập trắc nghiệm liên kết động từ-danh từ tiếng Anh dựa trên khối ngữ liệu BNC kết hợp hệ thống từ điển trực tuyến Wordnet để làm nguồn tài liệu học tập cho sinh viên tại Đại học ngoại ngữ Huế nói riêng, và những người học tiếng Anh ở Việt Nam như ngôn ngữ thứ 2 nói chung. Về Khối ngữ liệu Anh quốc (British National corpus), đây là một khối ngữ liệu lớn nhất với khoảng 100 triệu từ bao gồm khoảng 90% vị từ và cụm từ liên quan đến văn bản và khoảng 10% còn lại liên quan đến ngôn bản ở các thể loại khác nhau (Đào, 2007).

Chúng tôi chọn câu hỏi trắc nghiệm (BTTN) làm dạng bài tập vì BTTN là một trong những dạng phổ biến nhất trong các bài kiểm tra năng lực ngôn ngữ được phát triển chuyên nghiệp theo như Haladyna (2011) đã nêu trong nghiên cứu của mình. Chúng được sử dụng rộng rãi để đánh giá việc học ở cấp độ nhớ và hiểu (Coombe, Folse & Hubley, 2007). BTTN có ưu điểm là: (1) người học rất quen thuộc với định dạng này; (2) chúng có thể được sử dụng để đánh giá kiến thức ở các cấp độ khác nhau từ cơ bản đến nâng cao; (3) bài tập có thể được chấm điểm một cách tự động và khách quan; (4) nguồn ngân hàng bài tập rất lớn và luôn được cập nhật thường xuyên.

Vì vậy, trong nghiên cứu này, chúng tôi tập trung giải đáp ba câu hỏi chính như sau:

1. Làm thế nào để thiết kế xây dựng bài tập trắc nghiệm liên kết động từ - danh từ trong tiếng Anh bằng cách sử dụng khối ngữ liệu?
2. Xây dựng bài tập trắc nghiệm liên kết động từ - danh từ trong tiếng Anh có hiệu quả như thế nào khi sử dụng khối ngữ liệu?
3. Chất lượng bài tập trắc nghiệm liên kết động từ - danh từ trong tiếng Anh khi sử dụng khối ngữ liệu được đánh giá như thế nào?

2. Liên kết từ trong tiếng Anh

2.1 Khái niệm và các phương thức xác định

Kể từ khi khái niệm *liên kết từ* (Collocation) được đưa vào lý thuyết ngôn ngữ học, hình như đã thiếu tính cụ thể trong các định nghĩa về liên kết từ ở các phương pháp xử lý ngôn ngữ học (Bartsch, 2004). Thuật ngữ Liên kết từ có nhiều định nghĩa khác nhau và thường khá mơ hồ trong ngôn ngữ học và giảng dạy ngôn ngữ (Nesselhauf, 2005) nên dễ bị nhầm lẫn. Mặc dù thuật ngữ này được sử dụng đa dạng, có thể xác định dựa trên hai quan điểm chính: phương thức định hướng dựa vào thống kê (Herbst, 1996) và phương thức định hướng dựa vào cụm từ.

2.1.1 Phương thức định hướng thống kê

Phương thức định hướng thống kê bắt nguồn từ John R. Firth (1957) khi giới thiệu các thuật ngữ được sắp xếp theo thứ tự và khả năng sắp xếp thứ tự của từ, vì vậy liên kết từ là sự kết nối các từ với nhau. Có những liên kết từ là thành ngữ tượng hình, và những liên kết từ chỉ là các cụm từ tự do (ví dụ, “Quân đội Anh” và “Văn học Anh”) (Firth, 1957). Từ nhiều ví dụ của Firth về các cụm từ kết hợp, chúng ta thấy rằng việc sử dụng thuật ngữ này không bị hạn chế đối với sự kết hợp của hai từ.

Tuy vậy, Howarth (1996) lập luận rằng Firth không quan tâm đến việc xác định các tiểu thể loại của cụm từ dựa trên khả năng thay thế bị hạn chế bởi vì ý nghĩa của cụm từ, mà trước hết

là “một sự trừu tượng hóa ở cấp độ ngữ đoạn” (Firth, 1957, tr. 190-215). Đó là lý do tại sao một số nhà ngôn ngữ học coi các thuật ngữ mà Firth sử dụng để mô tả khái niệm theo sắp xếp thứ tự là mơ hồ và quá rộng. Ý tưởng của Firth đã được các nhà ngôn ngữ học phát triển khi cho rằng liên kết từ là sự liên kết đặc trưng của các mẫu từ. Halliday và Sinclair đã đưa ra các điều kiện nghiên cứu về liên kết từ của Firth về lý thuyết và phương pháp một cách có hệ thống vào những năm 1960. Do vậy, liên kết từ đã trở thành một trong những vấn đề chính của ngôn ngữ học ngữ liệu.

Halliday (1966) tuyên bố cần có một kho ngữ liệu khoảng 20 triệu từ để phân tích sự liên kết từ. Sinclair, một nhà ngôn ngữ học hiện đại thế hệ đầu tiên và là người sáng lập dự án COBUILD, bắt đầu thu thập dữ liệu và thực hiện các phân tích định lượng bằng máy tính vào những năm 1960. Một trong những tác động lớn của ông đối với lý thuyết ngôn ngữ là ông đã sử dụng phương pháp thống kê dựa trên ngữ liệu để tìm các liên kết từ, quan sát thấy rằng các liên kết từ được xác định bởi ngữ cảnh xung quanh. Theo Sinclair (1991), liên kết từ là sự xuất hiện của hai hoặc nhiều từ trong một khoảng cách ngắn trong một văn bản. Khoảng cách của các từ thông thường tối đa chỉ bốn từ. Ví dụ, từ “hard” nghĩa là “khó khăn, vất vả” nhưng khi xuất hiện trong các liên kết từ “hard luck”, “hard facts”, “hard evidence” thì lại mang nghĩa là *sự may rủi, sự thật khó tin, và bằng chứng khó tin*.

Từ các định nghĩa và khái niệm về liên kết từ và một số thuật ngữ kỹ thuật do Sinclair (1966) đề xuất, các nhà nghiên cứu đã sử dụng các kỹ thuật máy tính để đo lường tính khác biệt của cụm từ, trích xuất cụm từ một cách tự động, phát triển các chương trình và kỹ thuật phân tích cụ thể để thực hiện các nghiên cứu về cụm từ bổ sung và xây dựng từ điển dựa trên văn bản. Tần suất, được coi là tiêu chí tuyệt đối trong lĩnh vực tính toán, đã có ảnh hưởng lớn đến việc nghiên cứu cụm từ trong các lĩnh vực khác mà có thể chỉ ra cách sử dụng cụm từ kết hợp thực sự mà không cần dựa vào trực giác của người bản ngữ.

2.1.2 Phương thức định hướng dựa vào cụm từ

Khác với nghiên cứu theo định hướng thống kê nhằm xem xét hiện tượng ngôn ngữ ở cấp độ rộng mà không tập trung vào một loại cụm từ cụ thể nào, cách tiếp cận dựa trên cụm từ đưa ra một quan điểm chính xác và hạn chế hơn, trong đó các cụm từ trước hết được xem như là các biểu thức của ngôn ngữ. Cách tiếp cận này nhấn mạnh trạng thái ngôn ngữ của các liên kết từ, xem chúng là các cấu trúc cú pháp được xây dựng đúng (Seretan, 2008). Trạng thái ngôn ngữ của các liên kết từ được nêu rõ trong các định nghĩa như sau:

“đó là sự xuất hiện cùng lúc của hai hay nhiều mục từ vưng để hiện thực hóa các yếu tố cấu trúc trong một khuôn mẫu cú pháp nhất định” (Cowie, 1981, tr. 223-235)

“về mặt từ vựng và/hoặc ngữ dụng học, đó là sự xuất hiện lặp đi lặp của ít nhất hai mục từ vưng có quan hệ cú pháp trực tiếp với nhau” (Bartsch, 2004, tr.11-14)

Cowie là một trong những nhà ngôn ngữ quan trọng nhất theo cách tiếp cận định hướng dựa vào cụm từ, vì ông đã nỗ lực xác định các liên kết từ và phân loại kết hợp từ khác nhau một cách chính xác nhất (Nesselhauf, 2004, tr.223-242). Theo Cowie (1981, tr.223-235), liên kết từ là “sự xuất hiện đồng thời của hai hoặc nhiều mục từ vưng như là sự hiện thực hóa các yếu tố cấu trúc trong một mẫu cú pháp nhất định”. Ông nhấn mạnh tầm quan trọng của phạm vi từ để phân biệt các kết hợp mở và kết hợp bị hạn chế.

2.2 Các loại liên kết từ trong tiếng Anh

Với hai phương thức định hướng nêu trên, các liên kết từ trong tiếng Anh đã được phân loại dựa trên các thuộc tính và tiêu chí khác nhau. Benson, Benson và Ilson hay viết tắt là BBI (1997) phân loại các liên kết từ dựa trên ngữ pháp và từ vựng. Liên kết từ ngữ pháp chiếm ưu thế là danh từ, tính từ/phân từ, động từ, và một giới từ với cấu trúc ngữ pháp (BBI, 1997). Liên kết từ là một kiểu cấu trúc trong đó một động từ, danh từ, tính từ hoặc trạng từ tạo thành một kết nối mà sự xuất hiện của từ này có thể dự đoán được với sự xuất hiện của một từ khác. Các loại liên kết từ vựng khác nhau sẽ được liệt kê trong bảng 1 dưới đây:

Bảng 1. Phân loại liên kết từ của Benson, Benson và Ilson (BBI, 1997)

Thể loại	Ví dụ
Động từ + Danh từ/Đại từ/Phân từ	Compose music (soạn nhạc), set a record (lập kỷ lục), set an alarm (đặt đồng hồ báo thức)
Động từ + Danh từ	Take medicine (uống thuốc)
Danh từ + Động từ	Alarms sound (chuông báo thức reo)
Tính từ + Danh từ	Strong tea (trà đậm)
Danh từ + Danh từ	A school of fish (một đàn cá)
Trạng từ + Tính từ	sound asleep (ngủ say)
Động từ + Trạng từ	apologize humbly (khúm núm xin lỗi)

Hausmann (1984), một trong những nhà ngôn ngữ đầu tiên đề xuất cách xử lý có hệ thống đối với khái niệm sắp xếp trật tự từ trong ngôn ngữ học vào đầu những năm 1980 đã phân loại các tổ hợp từ theo hai đặc điểm *cố định* và *không cố định*. Như vậy, so với BBI (1997), ông đưa ra những hướng dẫn đơn giản tùy theo các kiểu quan hệ cú pháp và ngữ nghĩa khác nhau giữa các thành phần của một số cụm từ để giúp làm sáng tỏ bản chất của các cụm từ và mối quan hệ giữa các thành phần của chúng, gọi là phần gốc từ và phần liên kết của gốc từ.

Bảng 2. Các loại liên kết từ của Hausmann (1990)

Loại	Cấu trúc liên kết từ	Ví dụ
1	Động từ+Danh từ (làm tân ngữ)	tackle a problem (giải quyết vấn đề)
2	Tính từ+ Danh từ	weak tea (trà nhạt)
3	Danh từ (làm chủ ngữ)+Động từ	heart throbs (tim đập mạnh)
4	Danh từ + Danh từ	a hand of bananas (một nải chuối)
5	Trạng từ+ Tính từ	keenly aware (nhận thức sâu sắc)
6	Động từ+Trạng từ	hurt seriously (tổn thương nghiêm trọng)

Cowie (1981) nhấn mạnh tầm quan trọng của phạm vi từ để phân biệt các kết hợp mở và kết hợp có giới hạn và phân thành 4 loại kết hợp từ như sau:

- 1) Các kết hợp từ tự do: "giới hạn có thể được chỉ ra trên cơ sở ngữ nghĩa" và "tất cả các yếu tố của kết hợp từ được sử dụng theo nghĩa đen";
- 2) Các kết hợp từ có giới hạn: "có thể thay thế một số từ, nhưng bị hạn chế" và "ít nhất một thành phần không theo nghĩa đen và ít nhất một thành phần được sử dụng theo nghĩa đen";
- 3) Thành ngữ tượng hình: "hiếm khi có thể thay thế các yếu tố và "sự kết hợp có nghĩa bóng, nhưng vẫn giữ nguyên cách hiểu theo nghĩa đen;
- 4) Thành ngữ thuần túy: "không thể thay thế các yếu tố" và "sự kết hợp theo nghĩa bóng và không theo nghĩa đen".

Nesslhauf (2005) đề xuất chỉ sử dụng tính hoán đổi làm tiêu chí xác định cho sắp xếp trật tự từ. Do vậy có 3 loại kết hợp từ:

1. Kết hợp tự do (khi cả động từ và danh từ có thể sử dụng mà không bị hạn chế về tính hoán đổi, ví dụ, *read the book* = *đọc sách*);
2. Liên kết từ (danh từ có thể được sử dụng mà không bị hạn chế tùy theo nghĩa mà nó được sử dụng, còn động từ có thể bị hạn chế ở đối với một số danh từ làm tân ngữ, chẳng hạn, *do housework*, chứ không dùng *make housework*);
3. Thành ngữ (cả danh từ và động từ không có nghĩa riêng biệt mà chỉ dựa vào ngữ cảnh, ví dụ *kick the bucket* = *qua đời/chết* trong khi nghĩa của động từ và danh từ trong cụm từ này là “*đá cái xô*”).

2.3 Những thách thức của việc sử dụng liên kết từ tiếng Anh

Với sự đa dạng về ý nghĩa cũng như tầm quan trọng của việc sắp xếp thứ tự trong liên kết từ, có thể nói rằng sử dụng liên từ đúng và phù hợp với ngữ cảnh là một thách thức đối với người học tiếng Anh như một ngoại ngữ hoặc ngôn ngữ thứ hai (Hill, 2000; Smadja, 1989). Đã có những nghiên cứu về sự hiểu biết các liên kết từ của người học tiếng Anh chỉ ra rằng họ không có đủ kiến thức về liên kết từ và gặp nhiều khó khăn khi sử dụng liên kết từ trong môi trường giao tiếp bằng tiếng Anh, hay nói cách khác người học chưa nắm vững ý nghĩa và ngữ cảnh để sử dụng đúng các liên kết từ (Shei & Pain, 2000; Seretan, 2008). Ngoài ra do ngữ pháp và cấu trúc của liên kết từ giữa các ngôn ngữ không giống nhau, sự biến đổi ngữ nghĩa, sự thay đổi theo thời gian, vùng miền, văn phong, v.v. Hơn nữa, sự khác nhau về trật tự từ trong liên kết từ cũng là một thách thức đáng kể cho việc sử dụng collocations của người học tiếng Anh như ngôn ngữ thứ hai.

Thứ nhất, do nghĩa của liên kết từ khác với nghĩa của từng thành phần tạo nên liên kết đó nên người học dễ hiểu sai hoặc dịch sai trong quá trình chuyển ngữ. Một số ví dụ như “*touch and go*”, nếu dịch theo nghĩa đen “*sờ và đi*” hoàn toàn không liên quan đến nghĩa của liên kết từ này là “*mạo hiểm, trầm trọng, không chắc chắn*”; “*No end in sight*” có nghĩa “*không thấy hồi kết*” nếu phân tích nghĩa của từng từ trong liên kết này thì không thể hiểu được; “*Nail down*” nghĩa là “*thành công có được/chốt (hợp đồng), hiểu đầy đủ/nhận thông tin đầy đủ*”, chứ không liên quan gì đến bộ phận cơ thể “*móng tay*” (nail).

Thứ hai, do sự biến đổi và đồng nghĩa, một số liên kết từ có thể có các biến thể hoặc đồng nghĩa, ví dụ “*compose a song*” (“*sáng tác bài hát*”) chứ không phải “*create a song*”, “*do homework*” (“*làm bài tập*”) chứ không phải “*make homework*”, “*achieve my goal*” (“*đạt mục đích*”) chứ không phải “*carry out my goal*”, v.v.

Thứ ba, trật tự từ trong liên kết từ là một trong những khó khăn phổ biến đối với người học ngoại ngữ khi có sự khác biệt đáng kể về trật tự từ của tiếng mẹ đẻ (L1) và ngôn ngữ thứ 2 (L2) (Cao, 2022). Trật tự từ trong tiếng Anh và tiếng Việt thường trái ngược nhau, ví dụ Tính từ + Danh từ (tiếng Anh) chuyển thành Danh từ + Tính từ (tiếng Việt), ví dụ, “*a beautiful girl*” chuyển qua tiếng Việt không phải “*một đẹp cô gái*” mà là “*một cô gái đẹp*”, “*a lone wolf*” được chuyển thành “*con sói đơn độc, một người độc lập về tư tưởng*”.

2.4 Những nghiên cứu trước đây

Có nhiều nghiên cứu trước đây liên quan đến sử dụng liên kết từ, một phương pháp thường được sử dụng nhiều nhất là tìm gợi ý để đưa ra quyết định về liên kết từ trong các bài kiểm tra cụ thể, trong trường hợp đó, một liên kết từ được chọn trước làm trọng tâm nghiên cứu (Hussein, 1990; Bahns và Eldaw, 1993). Ví dụ, Hussein (1990) đã sử dụng một bài kiểm tra bốn mươi mục để đánh giá kiến thức về liên kết từ của 200 sinh viên đại học chuyên ngành tiếng Anh tại Đại học Yarmouk, Jordan. Kết quả cho thấy rất nhiều sinh viên chưa đạt yêu cầu do giáo viên quá chú trọng đến ngữ pháp trong cả giảng dạy và kiểm tra mà bỏ qua từ vựng, điều này dẫn đến người học chưa có kỹ năng đọc và đoán nghĩa của liên kết từ trong ngữ cảnh tốt được. Tương tự, khi kiểm tra kiến thức của những người học tiếng Anh trình độ cao cấp về các cụm động từ-danh từ qua bài tập dịch và điền vào chỗ trống, Bahns và Eldaw (1993) phát hiện ra rằng người học gặp nhiều vấn đề khi dịch các động từ hơn là dịch toàn bộ liên kết từ. Ngoài ra, phần lớn các lỗi gặp phải liên quan đến việc sắp xếp thứ tự trong liên kết từ do ảnh hưởng của tiếng mẹ đẻ.

Ngoài ra, nhiều nghiên cứu kiểm tra năng lực ngôn ngữ về liên kết từ của người học bằng cách trích xuất các liên kết từ trong các bài luận hoặc một số mẫu cụ thể hoặc từ kho ngữ liệu viết L2. Nesselhauf (2003) lần đầu tiên phân tích 32 bài luận được viết bởi những người học tiếng Anh trình độ cao, kiểm tra năng lực của họ bằng cách kết hợp động từ-danh từ. Lỗi chọn sai động từ thường xuyên xảy ra nhất. Ngoài ra, Nesselhauf quan sát thấy rằng sự chuyển đổi tiêu cực là một trong những yếu tố chính ảnh hưởng đến việc sử dụng đúng các cụm động từ-danh từ. Nesselhauf nhận thấy rằng các cụm động từ-danh từ không có từ tương đương trong bản dịch sang L2 trong khi L1 đã gây khó khăn hơn cho người học so với những cụm từ đồng nghĩa giữa L1 và L2. Vì vậy, việc giảng dạy các liên kết từ nên được tập trung nhiều hơn vào động từ (Nesselhauf, 2005).

Liu (1999) đã điều tra các lỗi liên kết từ trong bài kiểm tra cuối kỳ của 127 sinh viên đại học và 94 bài luận của họ. Phân tích lỗi cho thấy lỗi sắp xếp Động từ-Danh từ (động từ + danh từ/đại từ, động từ + mệnh đề) xuất hiện thường xuyên nhất. Liu (2002) xem xét các cách sắp xếp sai cụm động từ-danh từ trong các bài luận của người học được thu thập thông qua điều tra ngữ nghĩa từ vựng. Liu đồng thời chỉ ra rằng 87% các lỗi sắp xếp từ vựng (233/265) là do việc sử dụng sai cách sắp xếp động từ-danh từ và 93% trong số đó là do sử dụng sai các cụm động từ.

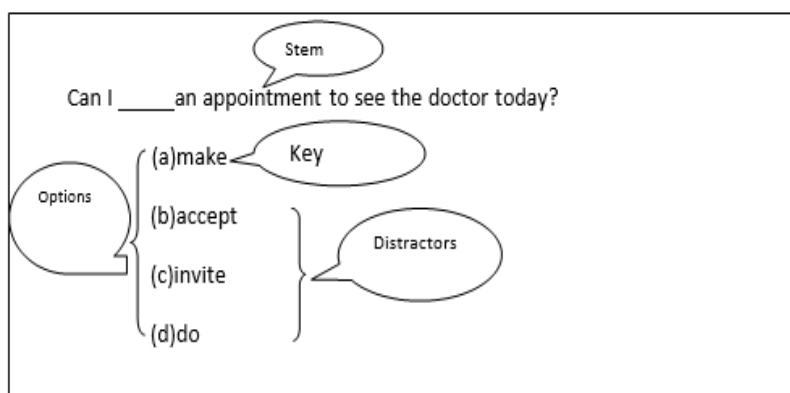
Ở Việt Nam, Cao và nhóm nghiên cứu (2022) đã tiến hành điều tra các đặc điểm về liên kết từ xuất hiện trong bộ sách giáo khoa tiếng Anh dành cho học sinh Việt Nam từ tiểu học đến trung học phổ thông bằng cách xác định các liên kết từ, tần suất xuất hiện và cách trình bày liên kết từ một cách phù hợp. Tổng cộng có 30.005 cụm từ động từ-danh từ và tính từ-danh từ đã được xác định, trong đó 10.077 liên kết động từ- danh từ. Nghiên cứu cho thấy rằng tần suất xuất hiện của liên kết từ xếp theo thứ tự tăng đều theo ba cấp độ lớp, nhưng lại không được phân bổ theo tỷ lệ phù hợp. Ngoài ra, chưa có sự nhất quán đối với tiêu chí đánh giá phương thức trình bày các liên kết từ trong ba bộ sách giáo khoa tiếng Anh.

3. Phương pháp nghiên cứu

3.1 Lập danh sách các liên kết từ để xây dựng bài tập trắc nghiệm

Trong nghiên cứu này, nhằm xây dựng các bài tập trắc nghiệm dựa trên khối ngữ liệu BNC, chúng tôi đã sử dụng phần mềm quản lý kho dữ liệu Sketch Engine để trích xuất tất cả liên kết từ mà không chỉ người Việt học tiếng Anh thường dễ phạm lỗi khi sử dụng (Thieu, 2017)

mà những người học tiếng Anh như một ngoại ngữ cũng thường gặp khó khăn trong việc xác định các kết hợp từ chính xác như người bản xứ (Vasiljevic, 2008). Dựa trên ví dụ của Isaacs (1994), đây là các phần chính để xây dựng một câu hỏi trắc nghiệm: 1/ xác định từ khóa; 2/ chọn từ gốc; 3/ xây dựng các thông tin nhiễu.



Hình 1. Các thành phần của câu hỏi trắc nghiệm (Isaacs, 1994)

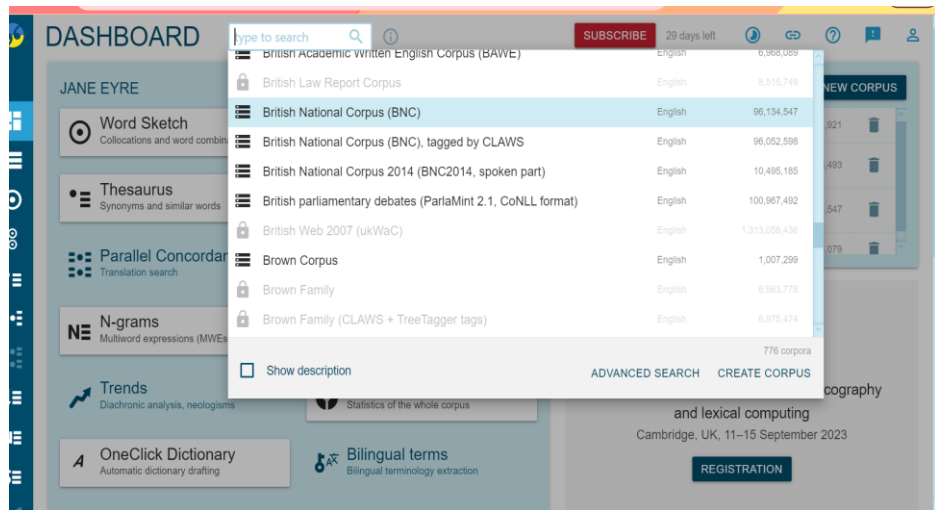
Hình 1 đã cho thấy 3 thành phần chính của câu hỏi trắc nghiệm: phần gốc(stem), từ khóa (key) và thông tin gây nhiễu (distractors). Phần gốc là câu hỏi mở đầu hoặc câu nói chưa hoàn chỉnh ở đầu mỗi mục và tiếp theo là các lựa chọn. Chúng tôi đã mượn cách tiếp cận phân loại lỗi sắp xếp thứ tự động từ-danh từ được đề xuất bởi Nesselhauf (2004) trong cuốn sách “Collocations in a Learner Corpus” trong nghiên cứu này để liệt kê những động từ cần xem xét liên kết với danh từ mà người học tiếng Anh có thể dễ bị mắc lỗi. Sau đó, chúng tôi sẽ dùng phần mềm Sketch Engine để tìm kiếm và trích xuất các câu có các liên kết từ đó trong khối ngữ liệu BNC.

3.2 Các bước thực hiện với phần mềm Sketch Engine để trích xuất các câu chứa các liên kết từ đã được liệt kê ở Mục 3.1 từ khối ngữ liệu BNC

Trong giai đoạn này, chúng tôi thực hiện các bước sau:

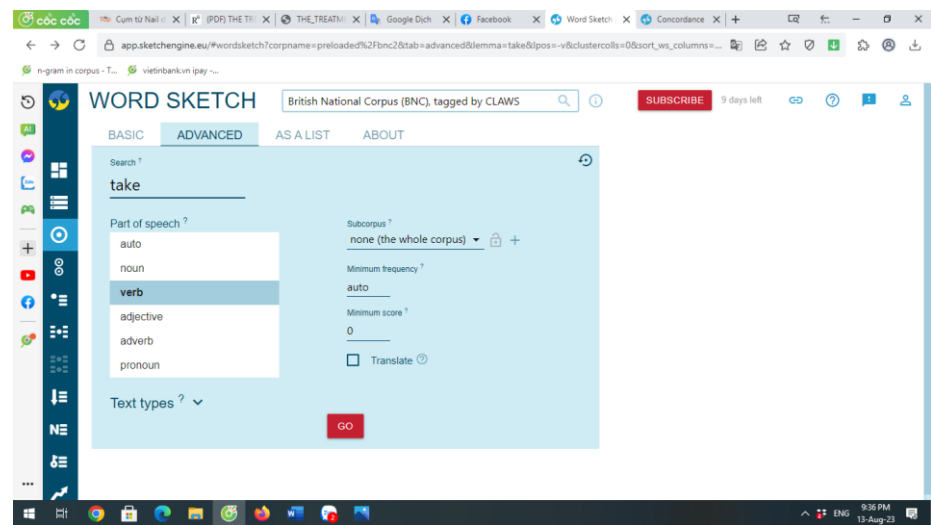
Bước 1: Chọn kho ngữ liệu để tìm liên kết từ

Trong danh mục nhiều khối ngữ liệu được xây dựng và cập nhật trên Sketch Engine, chúng tôi tiến hành chọn khối ngữ liệu British National Corpus (BNC). Sau khi đăng ký tài khoản và đăng nhập vào website: [https:// sketchengine.eu](https://sketchengine.eu), chúng tôi chọn khối ngữ liệu BNC.



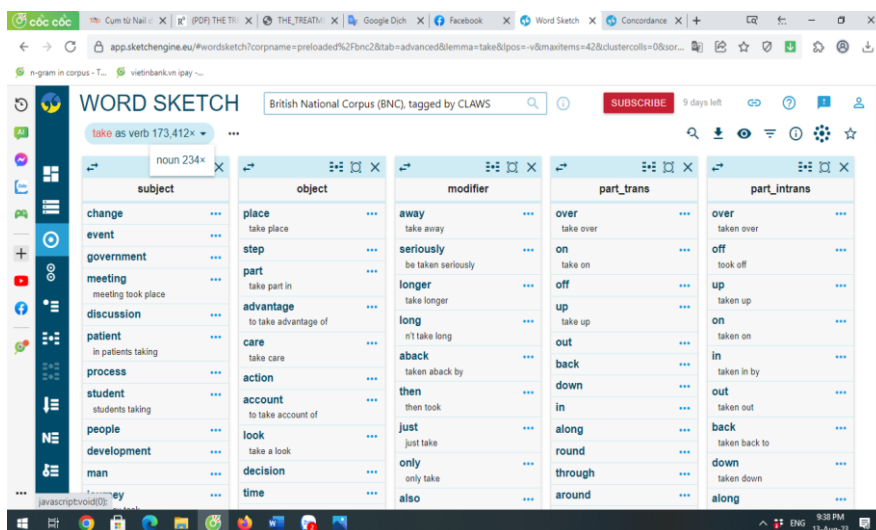
Hình 2. Chọn Khối ngữ liệu BNC với phần mềm Sketch Engine

Bước 2: Chọn Word Sketch và chọn từ khóa (tập trung vào động từ)



Hình 3. Chọn từ khóa để tìm kiếm kết từ

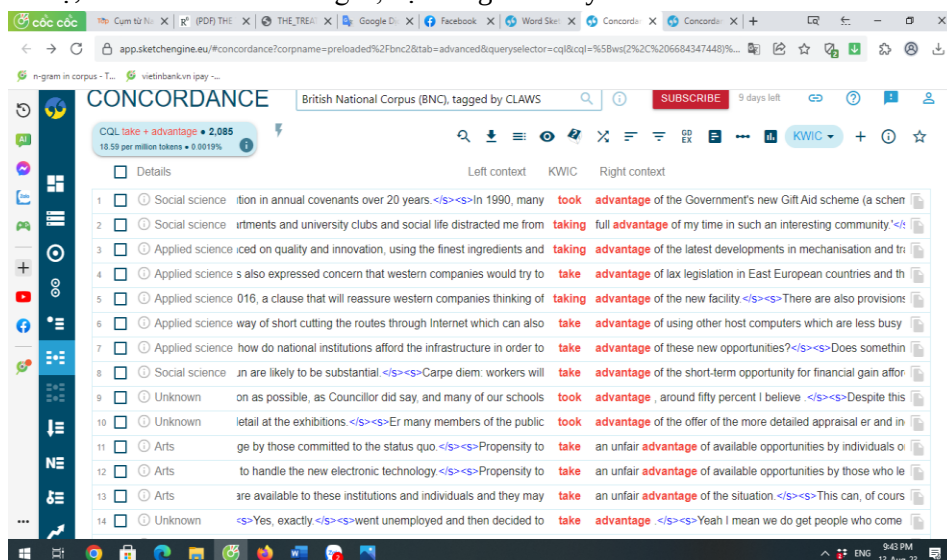
Chẳng hạn, với động từ Take, hệ thống sẽ cung cấp tất cả những từ đi kèm với “take” mang chức năng là chủ ngữ (subject), tân ngữ (object), bổ ngữ (modifier), và giới từ.



Hình 4. Minh họa những từ loại liên kết với từ khóa đã chọn

Bước 3: Nghiên cứu này chỉ tập trung vào liên kết từ Động từ+ Danh từ nên chúng tôi chọn mục Danh từ làm tân ngữ (Object).

Ví dụ, với danh từ “advantage”, hệ thống cho thấy 2.085 câu có chứa liên kết từ này.



Hình 5. Minh họa những câu chứa liên kết từ với từ khóa “Take”

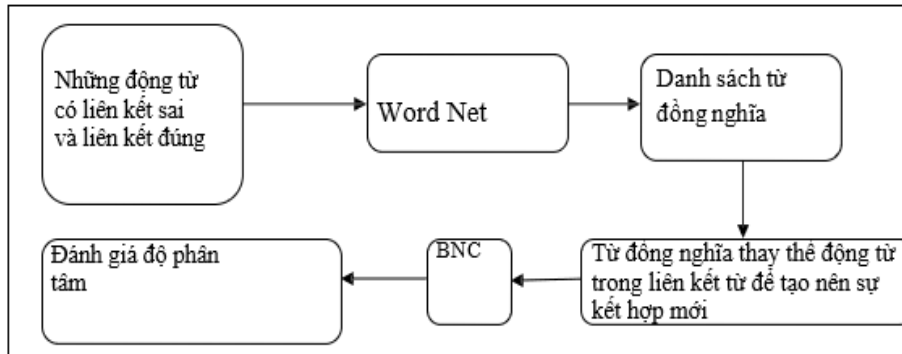
Bước 4: Trích xuất những câu chứa liên kết từ với động từ Take để xây dựng bài tập trắc nghiệm.

3.3 Xây dựng bài tập trắc nghiệm với các liên kết từ đã được trích xuất từ khối ngữ liệu BNC

Chúng tôi dựa trên yếu tố nhiễu phân tâm (distractor) làm cơ sở xây dựng những lựa chọn trong bài tập trắc nghiệm. Sau khi thử nghiệm với một số liên kết từ mà chúng tôi tự tạo các yếu tố gây nhiễu, chúng tôi phát hiện ra rằng các từ đồng nghĩa của các động từ sai được chọn có xu hướng trở thành yếu tố gây phân tâm tốt. Hiện tượng này thực chất tương ứng với lý thuyết về sự chuyển giao tiếng mẹ đẻ một cách tiêu cực như Liu (2002) đã mô tả. Do đó, dựa trên phân tích

về khả năng không thể thay thế của liên kết từ và nhiều phân tâm của ngôn ngữ thứ nhất (L1), chúng tôi quyết định truy xuất các từ đồng nghĩa của cả hai động từ trong liên kết động từ-danh từ có thể được người học chọn sai. Sau đó, để kiểm tra tính hợp lệ để trở thành những yếu tố gây phân tâm tốt, chúng tôi sẽ thay thế động từ trong cụm từ đồng nghĩa bằng các từ đồng nghĩa để tạo thành các kết hợp mới. Sau đó, chúng tôi sẽ kiểm tra xem các kết hợp mới có thực sự được sử dụng trong BNC hay không. Những động từ trong các liên kết không có lượt truy cập nào trong BNC sẽ được coi là những yếu tố gây phân tâm tốt.

Hình 6 dưới đây minh họa quá trình tạo nên độ phân tâm cho câu hỏi trắc nghiệm



Hình 6. Quá trình tạo yếu tố nhiễu phân tâm trong liên kết từ

Chúng tôi đã sử dụng Word Net (Miller, Beckwith, Fellbaum, Gross và Miller, 1993) để liệt kê các từ đồng nghĩa của các động từ. WordNet là một từ điển điện tử trong đó nghĩa của từ được cấu trúc theo thứ bậc. Danh từ, động từ, tính từ và trạng từ được nhóm lại thành các tập hợp từ đồng nghĩa nhận thức (synsets), mỗi từ thể hiện một khái niệm riêng biệt. Các tập hợp được liên kết với nhau bằng mối quan hệ khái niệm-ngữ nghĩa và từ vựng. Chúng tôi đã trích xuất các từ đồng nghĩa của các động từ mà người học thường gặp khó khăn khi sử dụng collocation. Những từ đồng nghĩa đó được thay thế bằng các động từ trong các cụm từ hoặc các danh từ đi kèm không xuất hiện trong BNC (nghĩa là sẽ không phải liên kết từ cần tìm) để xây dựng câu hỏi trắc nghiệm. Ví dụ một bài tập gồm 10 câu hỏi trắc nghiệm với liên kết động từ + danh từ của động từ “take” (xem Phụ lục 1).

3.4 Phương thức đánh giá BTTN có chứa các cụm từ liên kết (collocations)

Dựa trên các yếu tố đã nêu trên để xây dựng BTTN cho các cụm từ liên kết bao gồm những động từ/cụm từ chính, các yếu tố gây nhiễu để tạo sự phân tâm cho người chọn lựa câu trả lời đúng hay kiểm tra kiến thức về từ vựng của người học, độ dài của câu chứa cụm từ, và thời gian thực hiện việc xây dựng BTTN của giáo viên, chúng tôi đã chọn các tiêu chí sau đây để đánh giá: 1/Loại cụm từ liên kết; 2/ Cấu trúc ngữ pháp của phần chính (stem); 3/ Yếu tố phân tâm hay gây nhiễu thông tin; 4/ Độ dài của câu chứa cụm từ liên kết; 5/ Thời gian thực hiện xây dựng BTTN.

4. Đánh giá và thảo luận

Bằng cách sử dụng phần mềm Sketch Engine để trích xuất các liên kết từ (collocation) chứa động từ- danh từ từ kho ngữ liệu BNC chứa hơn 100 triệu từ, chúng tôi đã thực hiện việc xây dựng cho 500 BTTN. BNC là một nguồn tài liệu vô cùng phong phú, và luôn được cập nhật, nên việc xây dựng các bài kiểm tra gồm câu hỏi trắc nghiệm như thế này có thể được cải thiện bằng cách sử dụng văn bản xác thực, cập nhật về các chủ đề mà người học quan tâm.

4.1 Tính hiệu quả của phương pháp xây dựng bài tập trắc nghiệm về liên kết từ

Để có thể so sánh hiệu quả của phương pháp xây dựng bài tập liên kết từ này, chúng tôi đã thử nghiệm với nhóm 10 học viên Cao học ngành Ngôn ngữ Anh tại Đại học Ngoại ngữ Huế, trong đó có 5 người không sử dụng BNC và Sketch Engine, và 5 người chọn sử dụng nguồn dữ liệu và phần mềm này để thiết kết mỗi người 50 bài tập trắc nghiệm (BTTN) với 10 collocation gồm các động từ được chỉ ra. Chúng tôi đã nhận được 500 câu trắc nghiệm từ 2 nhóm tham gia với thời gian như sau:

Bảng 3: Tính hiệu quả của phương pháp xây dựng BTTN với BNC

Sử dụng Sketch Engine và BNC	Số lượng câu hỏi trắc nghiệm	Thời gian	Thời gian tính trung bình cho mỗi câu
Nhóm 1 (gồm 5 người không sử dụng)	250	1.000 phút	4 phút
Nhóm 2 (gồm 5 người có sử dụng) Engine và BNC)	250	150 phút	0,6 phút

4.2 Đánh giá về chất lượng bài tập

Để đánh giá chất lượng của các BTTN được xây dựng, chúng tôi đã tiến hành đánh giá kín: mỗi học viên sẽ nhận 50 BTTN ngẫu nhiên và không biết BTTN nào có sử dụng Sketch Engine và BNC dựa trên các tiêu chí đã nêu ở mục 3.4.

Kết quả cho thấy 85% BTTN có sử dụng BNC có thể được chấp nhận đáp ứng các tiêu chí trên, tuy nhiên vẫn còn 15% BTTN vẫn rất khó đoán đáp án đúng cho người học vì những lý do sau: 1) chưa có gợi ý rõ ràng trong phần chính (stem) để chọn câu trả lời đúng; 2) cấu trúc câu khó nên người học cảm thấy khó hiểu; 3) không đúng ngữ pháp và không phù hợp với mục đích rèn luyện từ vựng.

Điều này có thể dễ dàng thấy được lý do bởi vì trong khối ngữ liệu BNC, một số câu không đúng ngữ pháp vẫn được đưa vào từ nguồn tư liệu thực chưa được hiệu đính và chọn lựa phù hợp. Vì vậy, chúng ta cần đưa ra các tiêu chí để giúp lựa chọn những câu phù hợp. Đối với yếu tố phân tâm hay nhiễu thông tin, 120/500 BTTN trường hợp có nhiều hơn một lựa chọn gây khó khăn cho người học, ví dụ *take responsibility for* và *take charge of*; *take a rest* và *take a break*, vv.

5. Kết luận và đề xuất

Nghiên cứu này đã chỉ ra phương pháp xây dựng BTTN với liên kết từ (động từ- danh từ) tiếng Anh với việc sử dụng phần mềm Sketch Engine và khối ngữ liệu BNC. Đây là một điểm mới và sáng tạo so với cách xây dựng BTTN truyền thống khi chưa áp dụng các phần mềm quản lý kho ngữ liệu của ngành Ngôn ngữ học máy tính vào giảng dạy ngoại ngữ. Kết quả nghiên cứu này cho thấy rằng cách xây dựng BTTN với Sketch Engine và khối ngữ liệu BNC đã tiết kiệm thời gian gấp 7 lần so với phương pháp truyền thống (xem Bảng 3). Ngoài ra, do khối ngữ liệu luôn được cập nhật nên tính thực tiễn và xác thực với phong cách ngôn ngữ của người bản xứ được phát huy cao độ. Vì vậy, việc ứng dụng ngôn ngữ học khối liệu đóng vai trò quan trọng trong việc giảng dạy tiếng Anh nói riêng và các ngoại ngữ khác nói chung.

Ngoài ra, quá trình xây dựng BTTN này chỉ mới bước đầu thực nghiệm nên còn nhiều bất cập, chẳng hạn khối dữ liệu BNC quá lớn có thể làm người sử dụng không đủ thời gian khai thác hết. Ngoài ra, nhiều câu trong khối ngữ liệu vẫn chưa đúng ngữ pháp chuẩn và nội dung khó

hiểu nên cần phải hiệu chỉnh cho phù hợp. Hơn nữa, nghiên cứu chỉ mới tập trung vào các liên kết động từ-danh từ trong tiếng Anh có xu hướng gây khó khăn lớn nhất cho người học ngôn ngữ thứ hai, nhưng một số loại liên kết khác như tính từ-danh từ, động từ-trạng từ, danh từ-danh từ, v.v. sẽ không thể bị bỏ qua. Vì vậy, trong tương lai, chúng tôi sẽ tiếp tục điều tra các vấn đề khó khăn của người Việt học tiếng Anh để xây dựng các loại bài tập khác với các khối ngữ liệu khác như COCA.

References

- Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations?. *System*, 21(1), 101-114.
- Bartsch, S. (2004). *Structural and functional properties of collocations in English*. Tübingen: Gunter Narr Verlag, 11-14.
- Benson, M., Benson, E., & Ilson, R. (1997). *The BBI dictionary of English word combinations*. Amsterdam: John Benjamins.
- Cao, T.P.D. et al. (2022). The treatment of collocations in English textbooks for Vietnamese students. *TEFLIN*, 33(2).
- Coombe, C., Folse, K., & Hubley, N. (2007). A practical guide to assessing language learners. Ann Arbor: University of Michigan Press. *Corpus-based language studies: An advanced resource book* (By Tony McEnery, Richard Xiao, Yukio Tono).
- Cowie, A.P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics* 2, 3, 223-235.
- Đào, H.T. (2007). Ngôn ngữ học khối liệu (Corpus). *Tạp chí Ngôn ngữ và Đời sống*, 7(141), 9-13.
- Đặng, T.T.H (2021). Ứng dụng của ngôn ngữ học khối liệu trong nghiên cứu và giảng dạy ngoại ngữ qua ví dụ đối với tiếng Đức. *Tạp chí Khoa học Trường Đại học Cần Thơ*, 57(4C), 215-222. DOI:10.22144/ctu.jvn.2021.130.
- Firth, J.R. (1957). Modes of meaning. In F.R. Palmer (ed), *Papers in Linguistics* (pp. 190-215). London: Oxford University Press.
- Halliday, M.A.K. (1966). Lexis as a linguistic level. In C.E. Bazell, J.C. Catford, M.A.K. Halliday, & R.H. Robins (eds), *In Memory of J.R. Firth* (pp.148-162). London: Longman.
- Herbst, T. (1996). What are collocations: Sandy beaches or false teeth? *English Studies*, 77(4), 379-393.
- Howarth, P.A. (1996). Phraseology in English academic writing: *Lexicographica Series Maior*, 75. Tübingen: Max Niemeyer, 27-30.
- Haladynam, T.M. (2011). *Developing and validating multiple-choice test items*. Routledge. Taylor & Francis Group.
- Hausmann, F. (1990). Le dictionnaire de collocations. In F.J. Hausmann, H.E. Wiegand, & L. Zgusta (eds.), *Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Berlin: de Gruyter.
- Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching collocation* (pp.28-46). Hove, London: Language Teaching Publications.
- Hussein, R.F. (1990). Collocations: The missing link in vocabulary acquisition amongst EFL learners. *Papers and Studies in Contrastive Linguistics*, 26(26), 123-136.
- Isaacs, G. (1994). Multiple choice testing: A guide to the writing of multiple choice tests and to their analysis. *Higher Education Research and Development Society of Australia, Campbelltown, NSW*.
- Laufer, B., & Waldman, T. (2011). Verb-noun collocation in second language writing: A corpus analysis of learners' English. *Language learning*, 61(2), 647- 672.
- Lê, L.T., & Đỗ, X.D. (2021). Ứng dụng ngôn ngữ học ngữ liệu vào dạy học ngoại ngữ: Dạy học từ vựng tiếng Việt cho người nước ngoài qua ngữ liệu song ngữ Anh-Việt. *Tạp chí Khoa học Đại học Huế*:

Khoa học xã hội và nhân văn (ISSN 2588-1213), 130(6E), 67–81; DOI: 10.26459/hueunijssh.v130i6E.6386.

- Liu, C.P. (1999). An analysis of collocation errors in EFL writing. *The Proceedings of the English International Symposium on English Teaching* (pp. 483-494). Taipei: Crane Publishing Ltd.
- Liu, L.E. (2002). A corpus-based lexical semantic investigation. Master thesis Nattinger, J.R. (1980) a lexical phrase grammar for ESL. *TESOL Quarterly*, 14(3), 337-344.
- Miller, G., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1993). *Introduction to WordNet: An on-line lexical database*. Cambridge: MIT Press.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.
- Nesselhauf, N. (2005). Collocations in a learner corpus. *Studies in Corpus Linguistics* 14. Amsterdam: John Benjamins.
- Seretan, V. (2008). Collocation extraction based on Syntactic Parsing. Ph.D. Dissertation. University of Geneva.
- Sinclair, J. (1966). Beginning the study of lexis. In C.E. Bazell, J.C. Catford, M.A.K. Halliday, & Robins, R.H. (eds.), *In Memory of J.R. Firth*. Longman.
- Shei, C.C., & Pain, H. (2000). An ESL writer's collocational aid. *Computer Assisted Language Learning*, 13(2), 0958-8221.
- Sinclair, J. (1991). Corpus, Concordance, Collocation. Oxford University Press. Sinclair, J. (ed.), (1987). *Looking Up: An account of the cobuild project in lexical computing*. HarperCollins.
- Smadja, F. (1993). Retrieving collocations from text: Xtract. *Computational Linguistics*, 19(1), 143-177.
- Stephen, D. (2016). The impact of translation technologies on the process and product of translation. *International Journal of Communication*.
- Thiều, T.H.O. (2017). Giúp sinh viên tăng khả năng viết tiếng anh qua việc dạy ngữ và từ. *Tạp chí Giáo dục*, 420(2).
- Vasiljevic, Z. (2008). Teaching vocabulary to advanced Japanese students: A word association approach. *The East Asian Learner*, 4(1), 1-19.

<https://www.wordreference.com/englishcollocations/keep>.

CORPUS-BASED ENGLISH VERB+NOUN COLLOCATION TASK DESIGN

Abstract: The application of the corpus management software into the language teaching materials' development has become one of the active research tendencies in Computational Linguistics (CL). This paper presents the English verb - noun collocation task design based on BNC and WordNet which is used for English learners in Vietnam. With the Sketch Engine and BNC and Wordnet data, we have built around 500 multiple-choice questions for teaching English at the intermediate level. Then, we suggest using corpora in developing teaching materials of English grammar and expressing the NLP important role in teaching- learning English as well as other foreign languages in Vietnam.

Keywords: Corpus, noun-verb collocation, multiple choice questions