

BƯỚC ĐẦU KHAI THÁC KHO NGỮ LIỆU TIẾNG ANH CHUYÊN NGÀNH (ESP CORPUS): MỘT HƯỚNG MỚI CHO CÁCH TIẾP CẬN THEO NHU CẦU NGƯỜI HỌC

Th.S. Phan Thị Anh Nga

Th.S. Phạm Trần Minh Đức

Khoa Cơ Bản

TÓM TẮT

Ngôn ngữ học khối liệu (Corpus Linguistics) là một ngành khoa học trẻ và là giao điểm giữa khoa học ngôn ngữ và khoa học máy tính¹. Sự ra đời của ngôn ngữ học khối liệu đã làm nên “*cuộc cách mạng về nghiên cứu ngôn ngữ cũng như các ứng dụng của ngôn ngữ trong vòng vài thập niên qua*” (Hunston, 2002: tr.1). Tuy nhiên, việc khai thác và sử dụng các khối liệu ngôn ngữ trong giảng dạy ngoại ngữ nói chung và giảng dạy tiếng Anh chuyên ngành nói riêng ở Việt Nam hiện nay vẫn dừng ở con số khá khiêm tốn. Bài viết giới thiệu tóm lược một số nội dung chính liên quan đến ngôn ngữ học khối liệu; đề xuất một vài hoạt động trong lớp dạy tiếng dựa vào kho ngữ liệu có sẵn cùng với những gợi ý giúp giáo viên có thể tự khai thác khối liệu ngôn ngữ chuyên dụng phù hợp phục vụ cho giảng dạy và dịch thuật.

Từ khóa: *Ngôn ngữ học khối liệu, khối liệu, học hướng dữ liệu, tiếng Anh chuyên ngành.*

1. Đặt vấn đề

Trong những năm nửa cuối thế kỉ XX và đầu thế kỉ XXI, cùng với sự bùng nổ của các thành tựu về khoa học kĩ thuật và công nghệ cũng như sự phát triển vượt bậc của các thể hệ công nghệ máy tính trong ngôn ngữ học ứng dụng đã hình thành một xu hướng phát triển mới - **Ngôn ngữ học Corpus** (Ngôn ngữ học khối liệu). Theo đó, các khối liệu ngôn ngữ và phần mềm khối liệu trên máy tính ngày càng ảnh hưởng mạnh mẽ trong vai trò là nguồn tài liệu sư phạm phong phú, gọn nhẹ phục vụ đắc lực cho việc dạy và học ngoại ngữ².

Ở Việt Nam, khái niệm “ngôn ngữ học khối liệu” được giới thiệu qua các bài viết gắn liền với tên tuổi của **TS. Đào Hồng Thu**, xuất hiện lần đầu tiên trong ấn phẩm *Tạp chí Ngôn Ngữ và Đời Sống* tháng 3 năm 2007.

¹ **TS. Đào Hồng Thu** (2007). *Tạp Chí Ngôn Ngữ và Đời Sống* số 7(141), trang 9.

² **TS. Đào Hồng Thu** (2009). “Sử dụng khối liệu giảng dạy Tiếng Việt như một ngoại ngữ”. *Kỷ yếu Hội thảo quốc tế Nghiên cứu và giảng dạy ngôn ngữ, văn hóa Việt Nam – Trung Quốc ở Đông Á và Đông Nam Á* (lần thứ III), tháng 11/2009. NXB Đại học Quốc Gia Hà Nội, trang 555-560.

Cùng với các ứng dụng trong giảng dạy khác, việc học tiếng với sự trợ giúp của máy tính (Computer Assisted Language Learning: CALL); các hoạt động “học hướng dữ liệu” (Data - Driven Learning: DDL) (dẫn theo **Johns, 1986 & 1991**) và học dựa vào khối liệu sẽ là bước đột phá nhắm đến sự thay đổi từ “phương pháp giáo dục chú trọng vào trí nhớ” sang “phương pháp giáo dục chú trọng vào giải quyết vấn đề” (**Ken Watanabe, 2010: tr.7**).

Tại nhiều nước trên thế giới, các sản phẩm giáo dục dựa trên cơ sở khối liệu đã trở nên khá quen thuộc. Thế nhưng ở Việt Nam, khái niệm “khối liệu” vẫn còn là khái niệm hoàn toàn mới, đang ở trạng thái mở trong lĩnh vực giáo dục. Vì lẽ đó, các nghiên cứu về khai thác khối liệu trong dạy học tiếng Anh nói chung và tiếng Anh chuyên ngành nói riêng vẫn dừng lại ở con số hết sức ít ỏi.

Johns (2002) ghi nhận DDL học hướng dữ liệu là một quá trình trong đó cho phép “*người học tiếng tiếp cận dữ liệu trực tiếp giống như những nhà nghiên cứu ngôn ngữ học và mỗi học viên là một Sherlock Holmes*” (trang 108). Thực tế, đây là một phương pháp được nhiều người dạy và học ngoại ngữ thường tin dùng khi họ sử dụng kho ngữ liệu trực tuyến Internet để tìm kiếm các ngữ cảnh thực tế.

2. Các khái niệm cơ bản

Thuật ngữ “khối liệu” (corpus) hay được biết đến với tên gọi khác là “ngữ liệu” có nghĩa là “*kho dữ liệu, kho sưu tập tài liệu*” (theo Từ điển Anh-Việt, Trường Đại học Ngoại Ngữ, NXB Giáo Dục năm 2000, trang 368). Về tiếng Việt, các thuật ngữ “ngữ liệu”, “corpus” hay “corpus văn bản” có thể đặt ở vị trí tương đồng về ngữ nghĩa với thuật ngữ “khối liệu”. Từ “corpus” theo tiếng La-tinh có nghĩa là “any body of text” (khối văn bản bất kì – Đào Hồng Thu dịch).

Theo **Sinclair (1991)**, khối liệu là “*một khối các văn bản ngôn ngữ tự nhiên được chọn làm đại diện cho một trạng thái hay biến thể của một ngôn ngữ*”. Từ góc độ ngôn ngữ học, khối liệu là “*tập hợp các mẫu lớn mang tính đại diện cho một ngôn ngữ nhất định, được dùng làm nguồn tham khảo chuẩn cho ngôn ngữ đang được nghiên cứu*” (**Baker, 2006: tr.2**). Có thể mượn tượng về khối liệu như một cuốn sách tham khảo trong đó có chứa tất cả những từ ngữ minh họa được sử dụng trong ngữ cảnh cụ thể hoặc trích dẫn từ các tác phẩm của một tác giả cụ thể trong một loạt ngữ cảnh kèm theo mà từ đó xuất hiện. Nói cách khác, nếu từ điển là tập hợp các tri thức về ngôn ngữ, cách sử dụng từ cũng như ý nghĩa của từ thì khối liệu là tập hợp dữ liệu liên quan đến *cách viết, cách sử dụng từ đó trong thực tế*.

Trong một công trình nghiên cứu chung, **T. McEnery và A.Wilson (1996)** đã đưa ra định nghĩa cho khái niệm khối liệu như sau:

- a. (sử dụng tự do) khối liệu là văn bản bất kì;
- b. (sử dụng thường xuyên) khối liệu là văn bản điện tử;
- c. (sử dụng theo phong cách ngôn ngữ) khối liệu là văn bản điện tử, được tập hợp sao cho có sự hiện diện của tất cả các phong cách ngôn ngữ chức năng.

Như vậy về cơ bản, một tập hợp các văn bản bất kì đều có thể xem là một khối liệu. Ví dụ corpus về tuyển tập các tác phẩm của Chế Lan Viên, hay tuyển tập ca từ của Trịnh Công Sơn, một tập hợp câu hoặc văn bản có chứa tập hợp từ “Văn học Việt Nam” được đăng tải trên các báo, tạp chí v.v...

Thực ra, những khối liệu theo cách nghĩ tương tự đã xuất hiện từ lâu chủ yếu phục vụ mục đích nghiên cứu kinh thánh và được thực hiện hoàn toàn bằng phương pháp thủ công. Đơn cử, cuốn sách tham khảo kinh thánh bằng tiếng La-tinh vào thế kỉ XIII được coi là khối liệu hoàn chỉnh đầu tiên trên thế giới được xây dựng với sự giúp sức của hơn 500 mục sư (C.Tribble & G.Jones, 1997: tr.7). Hay một khối liệu khác gồm 11 triệu đơn vị từ và cụm từ bằng tiếng Đức được tạo lập bởi một học giả người Đức, Wilhelm Kading vào cuối những năm của thập niên 1890 (dẫn theo Sampson, G. & McCarthy, D., 2004). Thông thường, những khối liệu như vậy được xem là sản phẩm vô giá và duy nhất. Điều đó cũng đồng nghĩa với nguy cơ cao khối liệu dễ bị hư hỏng hay mất mát vì những lí do ngoài ý muốn. Điều thú vị là, cũng thuật ngữ “khối liệu” ấy nhưng nếu xét từ góc độ ngôn ngữ học khối liệu hiện đại thường được hiểu là ngôn ngữ học “khối liệu điện tử” (Sampson, G. & McCarthy, D., 2004: tr.1) thì thuật ngữ “khối liệu” khi được sử dụng trong ngôn ngữ học máy tính sẽ bao hàm nhiều ý nghĩa đặc trưng hơn nhiều so với định nghĩa đơn giản vừa nêu trên.

Leech (1991: tr.116) đưa ra một định nghĩa khá chặt chẽ về khối liệu điện tử như sau: *“những tập hợp các văn bản một cách có tổ chức: chúng thường được xây dựng với các mục đích cụ thể trước và được dùng làm đại diện cho một ngôn ngữ hay thể loại văn bản”*. Đồng quan điểm với Leech, Sinclair (1991) nhấn mạnh tính đại diện của khối liệu điện tử như sau: *“một tập hợp các phân đoạn của ngôn ngữ được lựa chọn và sắp xếp theo các tiêu chí ngôn ngữ học rõ ràng được sử dụng như một mẫu của ngôn ngữ đó”*.

Để dễ hiểu hơn, một kho ngữ liệu có thể tóm tắt theo một số cách diễn đạt khác như sau:

- (nghĩa rộng) bất cứ khối văn bản nào của ngôn ngữ tự nhiên;
- (nghĩa thông dụng) một khối văn bản máy đọc được;
- (nghĩa hẹp) một tập hợp nhất định các văn bản máy đọc được, là mẫu đại diện lớn nhất cho một ngôn ngữ hay các loại biến thể (variety)³.

³ Theo Phạm Hiến (2012). “Sử dụng kho ngữ liệu trong giảng dạy Tiếng Việt”. *Từ điển học & Bách Khoa thư*, số 1 (15).

Như vậy, về nguyên tắc bất cứ tập hợp nào lớn hơn một văn bản đều có thể gọi là một khối liệu. Theo tác giả **Phan Hiền (2012: XX)**, thuật ngữ “corpus” khi được sử dụng trong văn cảnh ngôn ngữ học hiện đại thường để định danh một cách chặt chẽ hơn lối định nghĩa giản đơn thông thường. Nói chính xác hơn, “ngôn ngữ học khối liệu”, nhìn từ góc độ của ngôn ngữ học hiện đại, “là một ngành nghiên cứu ngôn ngữ dựa trên mẫu lấy từ ngôn ngữ thật của đời sống hàng ngày được tập hợp dưới dạng các văn bản điện tử và được tiến hành nghiên cứu bằng cả phương pháp định tính và định lượng” (**Baker, 2006: tr.1&2**).

Theo đó, nếu nhìn khối liệu từ góc độ là cơ sở của Ngôn ngữ học khối liệu – khoa học nghiên cứu các phương pháp xây dựng và sử dụng khối liệu với sự hỗ trợ của công nghệ máy tính thì một khối liệu phải bao hàm các thuộc tính bất khả li sau (**Aarts, 1991; Sinclair, 1991; Knowles, 1996; McEnergy & Wilson, 1996; Biber và đồng sự, 1998; Tognini Bonelli, 2001**):

- tính nguyên bản (**authenticity**)
- tính đại diện (**representativeness**)
- tính mẫu (**sampling**)
- kích cỡ xác định (**finite size**)
- dạng thức điện tử, ở dạng đọc được trên máy tính (**machine-readable form**)
- tham chiếu chuẩn (**standard reference**)⁴

Như vậy, căn cứ vào bản chất và hoạt động ngôn ngữ của khối liệu có thể định nghĩa khối liệu là “tập hợp các dữ liệu tương đồng về mặt ngôn ngữ, được trình bày dưới dạng model văn bản điện tử, theo các cấu trúc nhất định và được sử dụng để giải quyết các vấn đề về ngôn ngữ trong các lĩnh vực khoa học cụ thể. Khối liệu có thể được sử dụng làm cẩm nang hướng dẫn và tìm kiếm thông tin hay công cụ biên soạn các thể loại từ điển khác nhau, và cuối cùng là phục vụ cho việc dạy học bản ngữ và ngoại ngữ cũng như dịch thuật. Trên cơ sở khối liệu, chúng ta có thể nhận được các dữ liệu về tần số hình thái từ, đơn vị từ vựng, phạm trù ngữ pháp; có thể theo dõi được các thay đổi về tần số và ngữ cảnh ở các thời điểm khác nhau” (**Đào Hồng Thu, 2009: tr.9**).

Johns (2002) ghi nhận DDL học hướng dữ liệu là một quá trình trong đó cho phép “người học tiếp cận dữ liệu trực tiếp giống như những nhà nghiên cứu ngôn ngữ học và mỗi học viên là một *Sherlock Holmes*” (trang 108).

⁴ Do khuôn khổ hạn hẹp của một bài báo, chúng tôi không thể trình bày hết ở đây, xin được trình bày kỹ những vấn đề liên quan ở các bài viết tiếp theo.

Cũng theo **TS. Đào Hồng Thu**, đối với các nhà nghiên cứu ngôn ngữ, việc sử dụng khối liệu cho phép họ tiết kiệm được rất nhiều thời gian và công sức. Chẳng hạn, chỉ cần gõ một từ bất kỳ vào khối liệu cần tìm kiếm, chúng ta có thể nhận được cả một danh mục liệt kê các trường hợp có sử dụng từ khóa với đầy đủ thông tin về nguồn gốc dữ liệu.

Tùy thuộc vào mục đích nghiên cứu mà khối liệu được chia thành những loại khác nhau như: khối liệu mẫu (sample corpora), khối liệu chủ (monitor corpus), khối liệu chú giải (annotated corpora), khối liệu song song (parallel corpora), khối liệu đối chiếu (comparable corpus), khối liệu tham khảo (reference corpus), khối liệu chuyên dụng (specialised corpus). Theo quan điểm của **Baker (2006)**, ông chia khối liệu thành 3 loại chính gồm:

- (1) *Khối liệu chuyên ngành*: dùng để nghiên cứu ngôn ngữ và phong cách ngôn ngữ thuộc các chuyên ngành khoa học cụ thể;
- (2) *Khối liệu lịch đại*: dùng để nghiên cứu ngôn ngữ và các biến thể ngôn ngữ trong một khoảng thời gian nhất định;
- (3) *Khối liệu tham khảo*: là khối liệu lớn đại diện cho một ngôn ngữ nhất định và giúp tìm ra nguồn dẫn của những văn bản cụ thể.

Khác với quan điểm của Baker, **Sinclair (1991)**⁵ lại phân chia khối liệu thành 2 loại chính đó là:

- (1) *Khối liệu mẫu* gồm nhiều đoạn trích ngắn có cùng phong cách ngôn ngữ được lựa chọn ngẫu nhiên, khoảng 1 triệu đơn vị từ và cụm từ sử dụng.
- (2) *Khối liệu tổng hợp* là loại khối liệu không có giới hạn cho phép nhà nghiên cứu truy cập với những mục đích sử dụng khác nhau và được xem như “ngân hàng ngôn ngữ”.

Tác giả bài viết này rất tâm đắc với phân loại các khối liệu chủ yếu của **TS. Đào Hồng Thu** (2009: tr.18), cụ thể:

- Theo kiểu dữ liệu, các khối liệu được chia thành khối liệu viết, khối liệu nói, khối liệu kết hợp;
- Theo ngôn ngữ văn bản, các khối liệu được chia thành khối liệu tiếng Việt, khối liệu tiếng Anh, khối liệu tiếng Nga;
- Theo đặc tính song song của văn bản, các khối liệu được chia thành khối liệu đơn ngữ, khối liệu song ngữ và khối liệu đa ngữ;
- Theo thể loại chuyên ngành, các khối liệu được chia thành khối liệu khoa học phổ biến, khối liệu không phổ biến, khối liệu ngôn ngữ học ứng dụng, khối liệu khoa học kỹ thuật và công nghệ;

⁵ **Hoàng Thị Ngọc Diễm (2012)**. “Khai thác khối liệu ngôn ngữ trong dạy và học tiếng Anh”. *Ngữ học toàn quốc 2011*. Hà Nội, trang 320.

- Theo tính chất chú giải, các khối liệu được chia thành khối liệu có chú giải và khối liệu không có chú giải;
- Theo chức năng và mục đích sử dụng, các khối liệu được chia thành khối liệu nghiên cứu, khối liệu minh họa, khối liệu tham khảo;
- Theo tính chất xã hội của văn bản, các khối liệu được chia thành khối liệu cộng đồng, khối liệu tác phẩm.

Theo tìm hiểu của chúng tôi, có hai kiểu khối liệu được sử dụng phổ biến và cho hiệu quả cao trong quá trình giảng dạy ngoại ngữ hiện nay. Đó là:

® *Khối liệu dành cho người học (Learner Corpus)*. Đây là công cụ cần thiết và quan trọng nhất đối với quá trình dạy và học ngoại ngữ, trong đó gồm các loại văn bản và ngôn bản bằng ngoại ngữ của người học được lưu trữ dưới dạng cơ sở dữ liệu và có thể được giảng viên hoặc chính người học phân tích trong quá trình dạy và học. Kiểu khối liệu này cho phép theo dõi quá trình học tập, phát hiện các lỗi kỹ năng và khoảng trống về kiến thức của người học.

® *Khối liệu song song (Parallel Corpus)*. Kiểu khối liệu này được sử dụng chủ yếu trong nghiên cứu khoa học và dịch thuật. Ngoài ra, khối liệu song song ngày càng được sử dụng hiệu quả trong dạy và học ngoại ngữ.

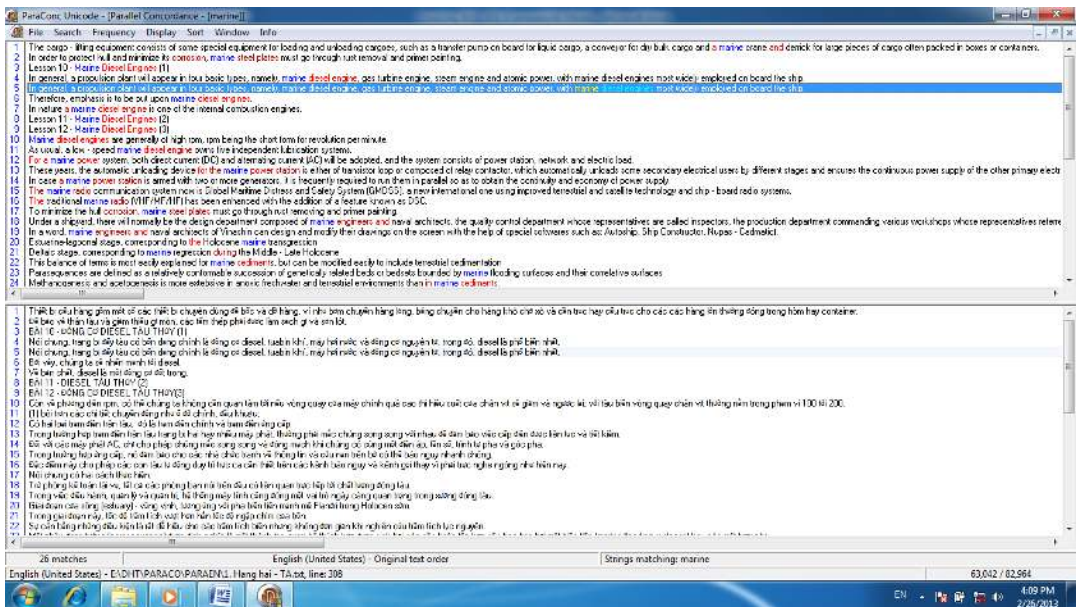
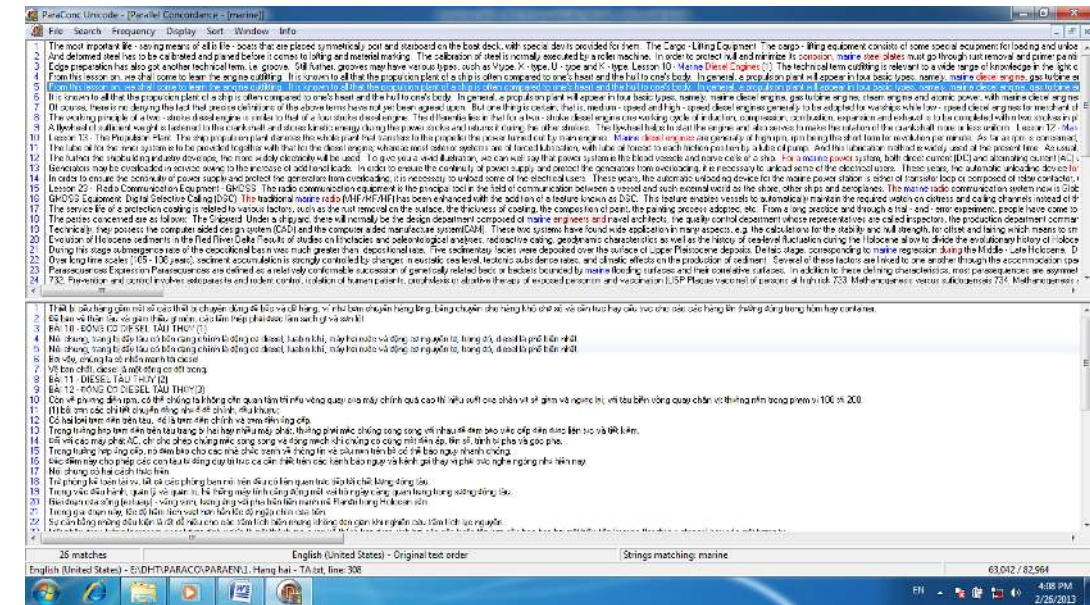
Trình độ ngoại ngữ của người học càng cao thì khả năng thực hiện tìm kiếm, phân tích các kết quả nhận được trong hệ thống danh mục từ (concordance) và cụm từ theo tần số sử dụng càng cao. Đồng thời, nội dung khối liệu ngôn ngữ ảnh hưởng trực tiếp đến hiệu quả tiếp nhận bài học và đóng vai trò quan trọng trong quá trình giảng dạy.

Một tác dụng đáng chú ý của công nghệ ngôn ngữ trong giảng dạy ngoại ngữ là khả năng sử dụng bảng danh mục (concordance) trên lớp học. Các loại bảng danh mục (concordance) có thể sử dụng để:

- Mô phỏng các ví dụ xác thực trong sử dụng các đơn vị từ vựng nhất định, các cụm từ, cấu trúc ngữ pháp, nội dung văn bản;
- Hình thành các nguyên tắc ngữ pháp và nguyên tắc sử dụng các đơn vị từ vựng;
- Mở rộng vốn từ điển (tìm kiếm nghĩa mới của các từ đã có trong từ điển, tương quan của từ và cụm từ với ngữ cảnh, nhập ngữ liệu vào từ điển một cách độc lập hoặc nhờ trợ giúp của giảng viên, so sánh các văn bản được tạo lập bởi người bản ngữ và người nước ngoài và v.v.).

3. Các khối liệu điển hình

Kể từ khi khối liệu đầu tiên ra đời vào năm 1961, hiện nay đã và đang có rất nhiều kiểu loại khối liệu được xây dựng và phát triển. Các khối liệu này đã và đang đóng vai trò quan trọng trong các lĩnh vực của ngôn ngữ học ứng dụng. Các khối liệu điển hình có thể đề cập đến là:



4. Khối liệu tiếng Anh chuyên ngành là gì?

Theo Leech (1997), corpora thường được coi là có hai vai trò chính trong việc học ngôn ngữ, cả hai đều đóng vai trò quan trọng trong giảng dạy tiếng Anh chuyên ngành dưới cả hai dạng tác động: trực tiếp và gián tiếp. Vai trò trực tiếp của khối liệu tiếng Anh chuyên ngành được xem như một công cụ su phạm hữu ích được khai thác chủ yếu bởi người học nhằm thỏa mãn nhu cầu học ngôn ngữ trong lớp học ngoại ngữ. Việc sử dụng gián tiếp của khối liệu thể hiện qua vai trò cung cấp dữ liệu và kiến thức nghiên cứu làm cơ sở cho các tài liệu sư phạm. Nghiên cứu khối liệu cho phép quan sát các mẫu lặp đi lặp lại với số lượng lớn dữ liệu. Trong một đánh giá quan trọng gần đây về lĩnh vực này, Granger (2009) lập luận rằng việc phân tích chặt chẽ kho khối liệu của người học có một vai trò quan trọng trong việc thiết kế các tài liệu sư phạm giảng dạy tiếng Anh chuyên ngành. Một trong tâm khác của nghiên cứu khối liệu là mô tả các đặc điểm cụ thể của các cụm từ liên quan đến diễn ngôn học thuật. Ở góc độ khác, Nesi (2015) trình bày một trường hợp khác về

biên dịch kho ngữ liệu nhỏ, được lấy mẫu thủ công, giàu thông tin theo ngữ cảnh để nghiên cứu, giảng dạy và học tập trong tiếng Anh chuyên ngành. Ngoài ra, cũng có thể khai thác khối liệu trực tuyến để tìm các tài liệu về các ngành học chuyên biệt, trong các miền cụ thể và với các đặc điểm ngôn ngữ liên quan đến các thể loại, chủ đề khác nhau.

Sự cần thiết của việc dạy-học theo khối liệu tiếng Anh chuyên ngành

Ngày càng có nhiều bằng chứng từ các nghiên cứu về khối liệu tiếng Anh chuyên ngành cho thấy tính hiệu quả cao trong áp dụng vào thực tế dạy học ngoại ngữ. Đối với những người quan tâm đến việc dạy và học ngoại ngữ thứ hai hoặc ngoại ngữ nói riêng, thông tin về sự phân bố các yếu tố và quá trình của một ngôn ngữ có thể ảnh hưởng đến phương pháp sư phạm theo một số cách. Theo quan điểm của Kennedy, trước hết nó có thể ảnh hưởng đến nội dung của việc giảng dạy ngôn ngữ bằng cách ảnh hưởng đến việc lựa chọn những gì sẽ dạy, trình tự của phương pháp sư phạm. Thứ hai, thông qua việc nâng cao nhận thức của giảng viên về ngôn ngữ và cách sử dụng ngôn ngữ, có thể cho thấy rằng khả năng xảy ra hay tần suất sử dụng là một thước đo quan trọng để đánh giá mức độ hữu ích của khối liệu ngôn ngữ chuyên ngành. Các nghiên cứu về Corpus của **Kenedy (2000)** chỉ rõ phương pháp giảng dạy ngôn ngữ bằng khối liệu có ảnh hưởng đến cách tiếp cận và đưa ra các kỹ thuật và chính sách khuyến khích việc tự truy cập và hướng dẫn cá nhân hóa thông qua tương tác với văn bản được phân tích xác thực từ cơ sở dữ liệu khối liệu chuyên ngành.

Corpus là một cơ sở dữ liệu khổng lồ bao gồm rất nhiều thông tin ngôn ngữ trong thế giới thực. Nó dùng để truy xuất máy tính và sử dụng cho các nghiên cứu chuyên biệt. Nó ngày càng đóng một vai trò quan trọng trong nghiên cứu và giảng dạy ngôn ngữ hiện đại với những ưu điểm độc đáo là dung lượng lớn, kho khối liệu thực tế không giới hạn cộng với truy xuất nhanh và chính xác. Cùng với sự phát triển của khối liệu, đây hứa hẹn sẽ là một xu hướng nhiều tiềm năng khi xây dựng kho khối liệu lớn, toàn diện cho tiếng Anh chuyên ngành. Tiếng Anh chuyên ngành khác với tiếng Anh cơ bản ở các khía cạnh sau: thứ nhất, nó mang tính chuyên biệt về cách sử dụng từ vựng, tiêu chuẩn nghiêm ngặt về hiệu suất cú pháp và cấu trúc đoạn văn logic đặc biệt bị hạn chế bởi hình thức văn bản cụ thể.

5. Một vài hoạt động dạy Tiếng Anh chuyên ngành dựa vào khối liệu ngôn ngữ

***Hoạt động 1:** Nhận diện nghĩa của thuật ngữ và cách sử dụng thuật ngữ đó qua ngữ cảnh.*

Hoạt động này có thể áp dụng cho các chuyên ngành có trình độ học viên khác nhau.

Bước 1: Giảng viên yêu cầu học viên tập hợp 10 thuật ngữ hoặc văn bản học thuật có thể bài đọc hoặc bài báo chuyên ngành và tạo thành một khối liệu nhỏ giống bộ bài tập (student portfolios). Sau đó chạy chương trình tìm kiếm dòng chỉ mục ngữ cảnh trên một khối liệu chuyên dụng trực tuyến có sẵn cho giảng viên chỉ định.

Bước 2: Cho học viên đọc các dòng chỉ mục đó và đoán nghĩa của thuật ngữ. Sau đó, tiến hành gộp các dòng chỉ mục có cùng nghĩa với nhau.

Bước 3: Đặt câu với thuật ngữ vừa học được.

***Hoạt động 2:** Tìm hiểu tần số sử dụng thuật ngữ.*

Bước 1: Cho học viên nhập 10 thuật ngữ hoặc văn bản học thuật vào chương trình và tạo bảng danh sách tần số các thuật ngữ sắp xếp theo trật tự ABC và theo tần số từ cao xuống thấp.

Bước 2: Đề nghị học viên đọc lướt danh sách tần số và viết lại các thuật ngữ không biết. Xác định

xem thuật ngữ mình không biết nằm ở nửa trên hay nửa dưới của danh sách tần số.

Lưu ý: Có thể thực hiện cả hai hoạt động trên đồng thời. Khi học viên không biết nghĩa của thuật ngữ trong *hoạt động 2* thì giảng viên có thể chủ động chuyển sang *hoạt động 1* để đoán nghĩa thuật ngữ trong ngữ cảnh.

Hoạt động 3: Thuật ngữ sử dụng trong các thể loại ngữ vực khác nhau. Để thực hiện được hoạt động này, giảng viên yêu cầu học viên chuẩn bị trước các văn bản học thuật hoặc các văn bản hội thoại. Hoạt động nên sử dụng khi học viên đã làm quen và thành thạo các thao tác khai thác khối liệu chuyên dụng để học thuật ngữ.

Chúng tôi đã tiến hành khảo sát sơ bộ nhu cầu này đối với trên 500 sinh viên trường Đại học Bách khoa Hà Nội, Đại học Ngoại ngữ và Đại học Khoa học xã hội và nhân văn - Đại học Quốc gia Hà Nội về nhu cầu sử dụng các sản phẩm công nghệ khối liệu ngôn ngữ như dịch máy (dịch tự động), các loại hình bài học, bài tập và từ điển dưới dạng số hóa. Kết quả khảo sát trên cho thấy đến gần **85%** số sinh viên các trường đại học nói trên có nhu cầu sử dụng các sản phẩm của công nghệ khối liệu với mục đích học tập và thực hành cho tương lai lập nghiệp.

Bước 1: Chia lớp thành 2 nhóm. Giảng viên đề nghị học viên nhóm thứ nhất nhập 10 văn bản học thuật và các học viên nhóm còn lại nhập 10 văn bản hội thoại vào chương trình. Các học viên có thể chọn các văn bản khác nhau trong cùng một thể loại.

Bước 2: Cho học viên tạo các danh sách tần số thuật ngữ của các văn bản mà họ đã chọn và sắp xếp chúng theo trật tự tần số tức thuật ngữ có tần số cao nhất sẽ đứng đầu danh sách.

Bước 3: Đề nghị học viên viết 10 thuật ngữ.

Bước 4: Cho các học viên so sánh danh sách thuật ngữ của họ với danh sách thuật ngữ của các học viên khác cùng nhóm xem có thuật ngữ nào trùng nhau và không trùng nhau. Sau đó, tập hợp lại thành một danh sách chung gồm 10 thuật ngữ có tần số cao nhất của toàn nhóm.

Bước 5: Cho học viên nhóm này so sánh danh sách thuật ngữ có tần số cao nhất của nhóm mình với danh sách thuật ngữ tần số cao nhất của nhóm kia. Thảo luận về sự khác nhau của danh sách thuật ngữ cùng nhóm và khác nhóm. Tìm danh sách thuật ngữ giống nhau ở cả hai nhóm.

Lưu ý: Khi học viên gặp thuật ngữ mới, có thể cho họ làm *hoạt động 1* để đoán nghĩa của thuật ngữ trong ngữ cảnh.

- **Những điểm nên lưu ý khi sử dụng khối liệu trong giảng dạy tiếng Anh chuyên ngành**

Theo **Reppen (2010, trang 43)** đề sử dụng khối liệu có hiệu quả cần lưu ý những điểm sau:

Giảng viên nên có ý tưởng rõ ràng về chủ điểm chuyên ngành muốn dạy, chọn những khối liệu chuyên dụng phù hợp nhất cho bài giảng đảm bảo hướng thực hiện là hoàn thiện và dễ thao tác, đảm bảo các minh họa tập trung vào chủ điểm muốn dạy, cung cấp nhiều cách tương tác với tài liệu. Nếu sử dụng máy tính nên có kế hoạch dự phòng trong trường hợp máy tính có vấn đề như in trước một số danh sách tần số, chỉ mục ngữ cảnh, hoặc danh sách đồng tệp từ để thay thế trong trường hợp trực trực hay có sự cố về kỹ thuật.

6. Kết luận

Việc sử dụng khối liệu ngôn ngữ trong giảng dạy tiếng là một hướng mở trong giảng dạy tiếng Anh chuyên ngành ở Việt Nam. Sử dụng khối liệu ngôn ngữ trong giảng dạy ngôn ngữ vẫn còn là cách tiếp cận mới nên trong quá trình giảng dạy, giảng viên cần chủ động có những điều chỉnh phù hợp với hoàn cảnh và điều kiện thiết bị khoa học kỹ thuật của đơn vị mình. Khối liệu ngôn ngữ dù là một sản phẩm công nghệ tiên tiến nhưng quá trình khai thác sử dụng trong bối cảnh Văn hóa-Xã hội và nền Giáo dục của Việt Nam cần vận dụng sáng tạo, linh hoạt, tránh rập khuôn nhằm đạt hiệu quả cao nhất trong cả lĩnh vực nghiên cứu và giảng dạy ngôn ngữ.

LỜI CẢM ƠN

Chúng tôi xin trân trọng cảm tạ và gửi lời tri ân đến **TS. ĐÀO HỒNG THU**, người đã truyền cảm hứng và tiếp sức cho chúng tôi trong suốt quá trình thực hiện nghiên cứu này. Chúng tôi xin chân thành cảm ơn **TS. Trần Thị Mai Đào** bởi thiện chí và tâm huyết bà giành cho chúng tôi trong quá trình tiến hành phân loại và tạo lập khối liệu. Chúng tôi gửi lời cảm ơn đến **TS. Hoàng Thị Ngọc Điểm**, Đại học Thái Nguyên đã chia sẻ tài liệu và kinh nghiệm khai thác khối liệu ngôn ngữ trong quá trình dạy và học Tiếng Anh.

TÀI LIỆU THAM KHẢO

Aarts, J. (1991). Intuition-based and observation-based grammars. In K. Aijmer, & B. Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, New York: Cambridge University Press.

Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.

Baker, P. (2006). *Using Corpora in Discourse Analysis*. London and New York: Continuum.

Đào Hồng Thu (2007). *Tap Chí Ngôn Ngữ và Đời Sống* số 7(141), trang 9.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge, England; New York: Cambridge University Press.

Đào Hồng Thu (2009). “Sử dụng khối liệu giảng dạy Tiếng Việt như một ngoại ngữ”. *Kỷ yếu Hội thảo quốc tế Nghiên cứu và giảng dạy ngôn ngữ, văn hóa Việt Nam – Trung Quốc ở Đông Á và Đông Nam Á* (lần thứ III), tháng 11/2009. NXB Đại học Quốc Gia Hà Nội, trang 555-560.

Granger, S. (2009). *The contribution of learner corpora to second language acquisition and foreign language teaching*. In K. Aijmer (ed.), *Corpora and Language Teaching*. 13 – 32. Amsterdam: John Benjamins.

Hyland, K. (2000). *Disciplinary Discourses: Social Interactions in Academic Writing*. Harlow, UK: Longman.

Hoàng Thị Ngọc Điểm (2012). “Khai thác khối liệu ngôn ngữ trong dạy và học tiếng Anh”. *Ngữ học toàn quốc 2011*. Hà Nội, trang 320.

Johns, T. (1991). “From printout to handout: Grammar and vocabulary learning in the context of data-driven learning”. *English Language Research Journal*, 4, p.27-46.

Johns, T. (1986). “Micro-concord: A language learner’s research tool”. *System*, 14(2), p.151-162.

Johns, T. (2002). Data-driven Learning. The perpetual Challenge. In B. Kettenmann & G. Marko (Eds.), *Teaching and Learning by doing Corpus Linguistics* (pp. 107-117): Amsterdam: Rodopi.

Kennedy, G. (2000). *An Introduction to Corpus Linguistics*. Beijing: Foreign Language Teaching and Research Press, p281-282.

Knowles, G. (1996). Corpora, databases and the organization of linguistic data. In J. Thomas, & M. Short (Eds.), *Using corpora for language research* (pp. 14-26). London: Longman.

Leech, G. (1991). The state of the art in corpus linguistics. In K. Aijmer, & B. Altenberg (Eds.), *English corpus linguistics* (pp. 8-29). London, New York: Longman.

Leech, G. (1997). *Teaching and language corpora: A convergence*. In A. Wichman, S. Fligelstone, T. McEnery and G. Knowles (eds.), *Teaching and Language Corpora*. London: Longman. 1 – 23.

McEnery, T. & Wilson, A. (1996a). *Corpus linguistics*. Edinburgh: Edinburgh University Press.

McEnery, T., & Wilson, A. (1996b). *Web-based course on corpus linguistics web pages to be used to supplement the book "Corpus linguistics"*. Retrieved February 1, 2005, from

<http://bowland-files.lancs.ac.uk/monkey/ihe/linguistics/contents.htm>

Nesi, H. (2015). *ESP corpus construction: a plea for a needs-driven approach*. ASp, volume 68:724. <http://dx.doi.org/10.4000/asp.4682>

Phạm Hiền (2012). “Sử dụng kho ngữ liệu trong giảng dạy Tiếng Việt”. *Từ điển học & Bách Khoa thư*, số 1 (15), tr.20-24.

Sampson, G. (1996). From central embedding to corpus. In J. Thomas, & M. Short (Eds.), *Using corpora for language research* (pp.14-26). London: Longman.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Ken Watanabe (2010). *Problem Solving 101*. Nhà Xuất Bản Trẻ

Tribble, C., & Jones, G. (1997). *Concordances in the classroom: using corpora. A resource guide for teachers*. London: Longman.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam, Philadelphia: John Benjamins Publishing Company.

