

## NGÔN NGỮ HỌC KHỐI LIỆU TRONG NGHIÊN CỨU DỊCH THUẬT

LÊ CHÍ HIỆU<sup>1</sup>  
QUÁCH THỊ BÌNH THO<sup>2</sup>

**Abstract:** The aim of this paper is to review some relevant factors related to the corpus linguistics and its using in translation studies. Firstly, we describe some definitions of corpus, its developments in linguistics, especially in context of Vietnam (it is called corpus linguistics) and some kinds of corpora such as parallel corpus, comparable corpus, etc... Then, we attempt to present some tools for exploiting the corpus called concordance such as Antconc, Concordance, etc... They are useful in using the electric corpora. Finally, we outline the implementation of corpora in translation studies and in SLT-Haui and FTU Hanoi.

**Keywords:** *Corpus, parallel corpus, compared corpus, using corpus in translation*

### 1. Mở đầu

Các nhà Ngôn ngữ học khối liệu cho rằng Sinclair là người đầu tiên khai mở với dự án Cobuild tại Trường Đại học Birminham khi ông đưa ra khái niệm như sau: “một tập hợp các văn bản ngôn ngữ tự nhiên, được lựa chọn nhằm đặc tính hóa dạng thức và tính đa dạng của ngôn ngữ” [19, tr.171]. Nói một cách khác, khối ngữ liệu là một tập hợp ngôn ngữ tự nhiên (văn bản hoặc giải mã điển ngôn) được tạo ra với một mục đích đặc biệt nào đó. Từ đó, có thể thấy rằng khối ngữ liệu là một tập hợp mẫu ngôn ngữ được thiết kế một cách có chọn lọc. Cho tới thập kỉ 90 của thế kỉ trước, cùng với sự phát triển của khoa học máy tính, khối ngữ liệu và ngôn ngữ học khối liệu đã gặt hái được những thành tựu quan trọng.

Ngày nay, các nhà ngôn ngữ học ngày càng sử dụng nhiều khối ngữ liệu trong nghiên cứu và ứng dụng dịch thuật. Nó rất hữu dụng để tìm hiểu thuật ngữ trong môi trường văn cảnh gốc, xây dựng các công cụ dịch thuật tự động và hỗ trợ giáo viên và sinh viên so sánh ngôn ngữ thông qua các khối ngữ liệu dịch và khối ngữ liệu so sánh. Đặc biệt, khối ngữ liệu rất cần cho học từ mới và tìm các kết hợp từ và thuật ngữ khoa học trong văn cảnh gốc.

Việc ứng dụng các khối ngữ liệu cho thực hành dịch thuật cho sinh viên chuyên ngành ngoại ngữ tại một số trường đại học chuyên ngoại ngữ là một nhu cầu thực tế (ví dụ tại Trường Ngoại ngữ - Du lịch, Trường Đại học Ngoại thương,...). Tuy nhiên, chưa nhiều trường đưa ngôn ngữ học khối liệu vào giảng dạy như là một học phần bắt buộc mặc dù đã có nhiều tọa đàm, hội thảo chuyên ngành nói về sự hữu dụng của nó trong học thuật. Nghiên cứu của chúng tôi góp phần vào những viên gạch xây dựng nền móng cho việc áp dụng khoa học khối ngữ liệu vào nghiên cứu và thực hành dịch thuật trong tương lai.

### 2. Tổng quan tình hình nghiên cứu

Trong ngôn ngữ học ứng dụng, việc sử dụng những mẫu văn bản thực phục vụ cho nghiên cứu dịch thuật không phải là vấn đề mới. Tuy nhiên, sử dụng khối ngữ liệu trong dịch thuật đã có những bước tiến đáng kể trong thời gian vừa qua. Sự phát triển của công nghệ máy tính và các công cụ điện tử giúp các nhà ngôn ngữ và sinh viên thu thập dữ liệu một cách nhanh chóng và hiệu quả hơn. Từ những năm 2000, các trường đại học tại châu Âu và Pháp đã đưa các khóa học về khối ngữ liệu vào chương trình học chính khóa dành cho sinh viên năm cuối và thạc sĩ. Ví dụ, Đại học Paris 7 phát triển một số công cụ hỗ trợ sinh viên sử dụng học tiếng Anh, tiếng Pháp, tiếng Trung Quốc,... Tại Đại học Putra Malaysia, các giảng viên đã xây dựng kho ngữ liệu EMAS phục vụ cho việc học và giảng dạy tiếng Anh như một ngoại ngữ. Bên cạnh đó, Đại học Malaya phát triển 2 kho ngữ liệu Macle (kho ngữ liệu tiếng Anh người học) dựa trên các bài luận văn của sinh viên và Comel (kho ngữ liệu tiếng Anh Malay) dựa trên các dự án và kho ngữ liệu nói. Tại Việt Nam, việc xây dựng một kho ngữ liệu quốc gia là cần thiết. Một số tác giả đã xây dựng khối ngữ liệu như Vncorpus, EBN corpus,... Tuy nhiên, việc này vẫn dừng lại ở những kết quả hạn chế. Chúng tôi đang cố gắng đưa nghiên cứu Ngôn ngữ học khối liệu vào cho sinh viên ngành ngôn ngữ nhằm phục vụ cho giảng dạy và học tập. Hiện nay, tại Trường Ngoại ngữ - Du lịch, Trường Đại học Công nghiệp Hà Nội (ĐHCNHN), chúng tôi đã đưa nội dung 2 học phần giảng dạy môn Ngôn ngữ học khối liệu và Ngôn ngữ máy tính vào chương trình cử nhân Ngôn ngữ học. Điều này giúp sinh viên tiếp cận ban đầu vào chuyên ngành ngôn ngữ học ứng dụng và ngôn ngữ học máy tính.

<sup>1,2</sup> Trường Đại học Công nghiệp Hà Nội



Khối ngữ liệu song song có thể được sử dụng để cung cấp thông tin trong dịch thuật và chứng minh các kết cấu tương đương giữa yếu tố từ vựng và cấu trúc trong ngôn ngữ nguồn và ngôn ngữ đích hoặc để nghiên cứu các hiện tượng dịch thuật [18]. Nó cũng có ích vì cung cấp dữ liệu cho các nhà ngôn ngữ học tâm lý khi khám phá sự khác nhau khi thụ đắc ngôn ngữ 1 và ngôn ngữ 2 [12].

Từ những năm 90 của thế kỷ trước, Sinclair đã có nhận xét ngắn gọn về việc sử dụng phương pháp nghiên cứu dựa trên khối ngữ liệu trong dịch thuật: "...các khối ngữ liệu sẽ có ảnh hưởng lớn tới dịch thuật trong tương lai..." [19]. Đây được coi là buổi bình minh của ứng dụng khối ngữ liệu trong dịch thuật. Sau đó, Baker [5] bắt đầu phân tích tính tương đồng và khác biệt trong các bản dịch thuật của cùng một ngôn ngữ do các nhà dịch thuật khác nhau xây dựng. Các nghiên cứu của bà đã gợi ý rất nhiều cho nghiên cứu dịch thuật và là nguồn cảm hứng cho các dự án sử dụng ảnh hưởng khối ngữ liệu trong dịch thuật. Từ đó, Laviosa [13] xây dựng English Comparable Corpus (ECC) - (Khối ngữ liệu so sánh tiếng Anh), là sự kết hợp của Translational English Corpus (TEC) - (Khối ngữ liệu dịch thuật tiếng Anh) và các văn bản lấy từ British National Corpus (BNC) - (Khối ngữ liệu quốc gia tiếng Anh). Tại Trung Quốc, Huang và Wan [9] sử dụng khối ngữ liệu so sánh BNC và Gigaword Chinese Corpus để giải thích các nguyên tắc của dịch thuật. Phân tích cho thấy rằng sự khác nhau về văn hoá trong ngôn ngữ nguồn và ngôn ngữ đích có thể được sử dụng để nhấn mạnh bản chất của dịch thuật.

### 3. Phương pháp nghiên cứu: sử dụng khối ngữ liệu trong dịch thuật

Đối với một dịch giả, khối ngữ liệu là một trong những nguồn thông tin ngôn ngữ quan trọng khi họ cần tìm các cấu trúc dịch thuật tương đương hoặc tìm cách xác định các chức năng dịch thuật. Trước đây, hầu hết các dịch giả chủ yếu dùng từ điển và tìm các thông tin từ vựng phục vụ cho công tác dịch thuật. Chính vì vậy, Bowker và Pearson [7] đã liệt kê 5 vấn đề mà các dịch giả gặp phải khi sử dụng từ điển, đặc biệt đối với các văn bản ngôn ngữ chuyên ngành (LSP):

- Chưa hoàn chỉnh: để viết một cuốn từ điển đòi hỏi rất nhiều thời gian. Vì vậy, trong một số trường hợp, nó không phản ánh được hiện thực của nhận thức và ngôn ngữ.
- Dung lượng: phần lớn các từ điển trước đây được in ấn bằng khổ giấy.
- Thiếu thông tin ngữ cảnh.
- Thiếu thông tin tần suất: các dịch giả thường muốn biết tần số sử dụng của từ.
- Khó khăn trong việc tìm kiếm thông tin liên quan: đối với một cuốn từ điển truyền thống thì làm như thế nào để tìm kiếm các từ vựng đang ở đâu.

Sau đó, Kubker [17] tìm ra mối liên hệ giữa khối ngữ liệu và dịch thuật có thể nhìn dưới hai góc độ: (1) Sử dụng khối ngữ liệu để dịch thuật: tức là sử dụng khối ngữ liệu như là một công cụ để tìm các thông tin ngôn ngữ trong tiến trình dịch. (2) Dịch thuật sử dụng khối ngữ liệu: tức là quá trình nghiên cứu dịch thuật có sử dụng khối ngữ liệu.

Baker xuất bản bài báo với tiêu đề: "Ngôn ngữ học khối liệu và nghiên cứu dịch thuật: ngụ ý và ứng dụng" nhằm tôn vinh sự nghiệp của nhà ngôn ngữ học Sinclair. Bà dự báo: "Sự tiện dụng của khối ngữ liệu của ngôn ngữ nguồn và ngôn ngữ đích sẽ giúp các dịch giả khám phá bản chất của văn bản dịch" [5]. Hai năm sau đó, trong bài báo "Khối ngữ liệu trong nghiên cứu dịch thuật: quan điểm và những gợi ý cho nghiên cứu tương lai", bà đã phát triển ý tưởng bằng cách xây dựng các khối ngữ liệu song song, dịch thuật, đa ngữ. Đây được coi là buổi bình minh của ứng dụng khối ngữ liệu trong nghiên cứu dịch thuật.

Thực tế là ứng dụng khối ngữ liệu trong dịch thuật đang trở thành một hướng nghiên cứu mới liên quan tới thực hành và đào tạo dịch giả liên môn, trực tiếp tác động tới phân tích đối sánh, ngôn ngữ chuyên ngành, thuật ngữ, từ vựng học và ngôn ngữ máy tính. Gavioli và Zanetti đã thực hiện nghiên cứu khi các dịch giả đã thiết kế, nén dữ liệu và phân tích khối ngữ liệu so sánh Anh - Ý trong một bài báo về nghiên cứu y khoa. Hatim ca ngợi dịch thuật dựa trên khối ngữ liệu thực sự là một làn gió mới trong nghiên cứu dịch thuật [11]. Baker sử dụng thuật ngữ "*khối ngữ liệu đa ngữ*" (*multilingual corpus*) để chỉ "tập hợp hai hoặc nhiều khối ngữ liệu đơn ngữ bằng nhiều ngôn ngữ khác nhau..." [5].

Tuy nhiên, nếu cho rằng khối ngữ liệu đa ngữ là tập hợp các văn bản chưa được dịch được sử dụng cho nghiên cứu đối chiếu ngôn ngữ và từ vựng song ngữ thì Baker lại nhìn thấy tính hữu dụng của khối ngữ liệu đa ngữ trong dịch thuật: "một cách thật tự nhiên khối ngữ liệu đa ngữ cho thấy từ được sử dụng một cách tự nhiên nhất trong ngôn ngữ A và ngôn ngữ B" [5]. Thực tế là chúng cung cấp những kĩ thuật hữu ích cho thống kê và kết hợp từ trong một số lượng vô hạn văn bản [5]. Vì vậy, khối ngữ liệu và các phần mềm xử lý văn bản cung cấp cho dịch giả một công cụ mạnh mẽ nghiên cứu dịch thuật trong trạng thái nguyên bản nhất.

Baker đề xuất sử dụng khối ngữ liệu so sánh trong nghiên cứu dịch thuật dựa trên tuyển tập các văn bản dịch từ một hoặc nhiều ngôn ngữ sang tiếng Anh. Bà cho rằng việc sử dụng các văn bản dịch trong dịch thuật sẽ làm sáng tỏ hơn, đỡ nhầm lẫn hơn và tương thích ngữ pháp hơn. Thực tế cho thấy các khối ngữ liệu so sánh cho phép dịch thuật mạnh mẽ hơn [20].

NNHKL sử dụng phương pháp nghiên cứu tiếp cận dựa trên ngữ liệu (Corpus-based Approach-PPTCNL) để nghiên cứu ngôn ngữ dựa trên các tập hợp (*collections*) của ngôn ngữ thực (*real life*) lưu trữ trong các ngữ liệu lớn (*corpora*). Nói một cách đơn giản là phương pháp này nghiên cứu các hiện tượng ngôn ngữ thông qua các ngữ cảnh thực tế nhất nhờ công cụ hỗ trợ là các dữ liệu điện tử lớn và các phần mềm phân tách từ do những nhà ngôn ngữ học máy tính phát triển. Phương pháp tiếp cận này có ích cho giảng dạy ngôn ngữ, học ngoại ngữ, làm từ điển, phân tích ngữ pháp, phân tích ngữ cảnh,... nhờ vào cách soi chiếu chúng trong các văn cảnh (context) gốc với số lượng không hạn chế. PPTCNL dựa trên các nguyên tắc sau: (1) Ngữ liệu lớn (Data-Driven): việc áp dụng phương pháp hoàn toàn dựa vào ngữ liệu chứ không dựa vào tổng hợp lý thuyết. (2) Ngôn ngữ gốc (Authentic language): PPTCNL cung cấp các ngữ liệu nguyên bản, thể hiện cách sử dụng một cách thực tế nhất. (3) Phân tích định lượng và định tính (Quantity and Quality Analysis): các công cụ thống kê góp phần phân tích số liệu và đặc điểm hiện tượng ngôn ngữ.

Tại Việt Nam, chúng tôi đang cố gắng đưa NNHKL theo hướng *tiếp cận dựa trên khối ngữ liệu* vào học tập và nghiên cứu dịch thuật cho sinh viên năm cuối chuyên ngành ngoại ngữ và ngôn ngữ. Tại Trường Ngoại ngữ - Du lịch, ĐHCNHN, NNHKL trở thành một học phần bắt buộc đối với sinh viên năm thứ 3 chuyên ngành ngôn ngữ. Đây là học phần giúp các em làm quen với NNHKL và chuẩn bị cho học phần ngôn ngữ học máy tính ở năm thứ 4. Một trong những ứng dụng quan trọng của nó là thực hành trong dịch thuật với ngôn ngữ nguồn là tiếng Anh và ngôn ngữ đích là tiếng Việt và ngược lại. Bên cạnh đó, chúng tôi đã đưa thí điểm việc ứng dụng NNHKL vào dịch thuật cho sinh viên khoa Pháp năm thứ 4 tại Trường Đại học Ngoại thương. Các em sẽ học cách xây dựng các khối ngữ liệu chuyên ngành dịch thuật và thực hành trong các văn bản thương mại chuyên ngành. Mặc dù, việc tiếp cận còn mới nhưng đã đem lại những phản hồi tích cực từ phía sinh viên khi họ được tiếp cận một phương pháp dịch thuật mới bên cạnh các phương pháp dịch truyền thống đã tồn tại từ hàng chục năm nay trong các trường ngoại ngữ.

#### **4. Kết quả nghiên cứu: khối ngữ liệu trong nghề dịch thuật**

Thực tế là hầu hết các dịch giả phải dùng máy tính trong công việc dịch thuật của họ bởi vì các bản dịch hiện nay chủ yếu dưới dạng văn bản điện tử. Vì vậy, công việc dịch thuật luôn diễn ra trong môi trường có sự trợ giúp của máy tính (*computer-assisted translation - CAT*). Hiện nay, các phần mềm dịch thuật đang góp phần thay đổi nghề dịch vì nó đòi hỏi các kỹ năng đặc biệt và kiến thức dịch. Nó giúp dịch giả tiết kiệm thời gian, đảm bảo tính chính xác và hiệu suất cao. Như vậy, các dữ liệu bộ nhớ dịch thuật được hiểu như là một khối ngữ liệu so sánh mà trong đó các câu được dịch sẽ giống hệt với các câu trong văn bản nguồn. Trong trường hợp này, các công cụ dịch được thực hiện qua quá trình quét scan và dịch giả có nhiệm vụ kiểm tra lại các câu đã được dịch. Bộ nhớ dịch có thể trích xuất các chỉ mục ngữ cảnh song song (*parallel concordances*) từ các từ hoặc kết hợp từ trong văn bản nguồn.

Cuối cùng là nghề dịch đòi hỏi dịch giả phải có kỹ năng và kiến thức thực hành về ngôn ngữ học ngữ liệu. Họ phải có khả năng xây dựng được các khối ngữ liệu đơn ngữ, song ngữ và sử dụng thành thạo các chỉ mục ngữ cảnh (*concordance*) để tìm kiếm thông tin của từ, liên kết từ, thuật ngữ trong cả ngôn ngữ nguồn và ngôn ngữ đích. Đặc biệt là sử dụng các khối ngữ liệu song song để giải quyết các vấn đề về dịch thuật. Dịch giả có thể xây dựng cho mình các khối ngữ liệu nhỏ và phân tích các tài liệu bằng các công cụ tìm kiếm tự thiết kế (*ad hoc programs*) hoặc trực tuyến (*online services*) như Antconc, Wordsmith và các công cụ khác, v.v.

Tại Trường Ngoại ngữ - Du lịch, ĐHCNHN đã đưa NNHKL và Ngôn ngữ học máy tính vào chương trình học năm thứ 3 và thứ 4 (khoảng 80 sinh viên). Đây là một định hướng phục vụ cho sinh viên sau này có thể học lên cao hơn và làm việc trong lĩnh vực ngôn ngữ máy tính, một ứng dụng của ngôn ngữ học ứng dụng. Việc ứng dụng các kho ngữ liệu trong dịch thuật là một ứng dụng của ngôn ngữ học khối liệu. Tuy nhiên, đối tượng của sinh viên ngành ngôn ngữ học không phải là dịch thuật nên việc ứng dụng trong dịch thuật chỉ dừng lại ở mức độ giới thiệu và ứng dụng ở mức hạn chế.

Tại Đại học Ngoại thương, sinh viên chuyên ngành Ngôn ngữ Pháp được thực hành các môn dịch từ năm thứ 3 trình độ B2 và C1. Phương pháp dịch sử dụng chủ yếu cho sinh viên vẫn là phương pháp truyền thống như dịch tóm tắt, dịch song song,... Chúng tôi đã sử dụng phương pháp dịch dựa trên khối ngữ liệu (*corpu-based approach*) áp dụng vào các lớp dạy dịch năm thứ 4 (với đối tượng sinh viên của 2 lớp, khoảng 80 sinh viên).

Đầu tiên, sinh viên nhóm 1 xây dựng một kho ngữ liệu người học bằng 2 ngôn ngữ Anh - Pháp. Đây là một kho ngữ liệu dịch (compared corpus) với khối ngữ liệu là các văn bản trong lĩnh vực thương mại. Sau đó, các sinh viên có thể dùng phần mềm Antconc hoặc Wordsmith để khai thác các khối ngữ liệu dịch thuật để tìm hiểu các thuật ngữ và các đoạn dịch bằng 2 ngôn ngữ này.

Nhóm thứ 2 sử dụng các kho ngữ liệu online Tradoit (tham khảo tại [tradoit.com](http://tradoit.com)). Đây là kho ngữ liệu Anh - Pháp đa lĩnh vực giúp các dịch giả có thể sử dụng online mà không cần xây dựng cho mình một kho ngữ liệu riêng của người học. Sinh viên có thể sử dụng kho ngữ liệu 2 chiều Anh - Pháp và Pháp - Anh. Thông qua đó, sinh viên có thể so sánh các thuật ngữ hay các bản dịch 2 chiều. Từ đó, rút ra các kinh nghiệm dịch thực tiễn bằng 2 ngôn ngữ. Chúng ta cùng quan sát kết quả một chỉ mục Anh - Pháp mà chúng tôi trích dẫn từ nguồn ngữ liệu Tradoit, một trang web khối ngữ liệu so sánh. Đây là trường hợp dịch từ “thanks to” từ tiếng Anh sang tiếng Pháp. Với một khối ngữ liệu so sánh 200 triệu từ thì đây là một công cụ vô cùng mạnh mẽ cho dịch thuật và tìm kiếm thuật ngữ.

Chúng ta cùng quan sát một ví dụ khác trên khối ngữ liệu dịch Pháp - Anh đối với từ “grâce à” (Nhờ vào) (Ảnh 3).

Quan sát cả hai ví dụ, sinh viên có thể thấy dù hai từ cùng nghĩa “thanks to” và “grâce à” nhưng bản dịch có sự khác biệt tương đối rõ ràng. Nhiệm vụ của sinh viên là phải đối chiếu, nhận xét và tìm thấy những bản dịch hợp lý nhất. Như vậy, bên cạnh công cụ dịch thuật mạnh mẽ thì chúng ta không thể phủ nhận vai trò của chính các dịch giả.

### 5. Thảo luận và hướng nghiên cứu trong tương lai

NNHKL không phải là tiếp nhận dữ liệu từ máy tính mà đó là quá trình nghiên cứu, phân tích dữ liệu từ các khối ngữ liệu. Nhiệm vụ của nó không phải là tìm dữ liệu mà là quá trình phân tích nó trong văn bản gốc. Chính vì vậy, máy tính là công cụ hữu hiệu trong quá trình phân tích đó. Vậy làm thế nào để giúp sinh viên sử dụng tốt các công cụ này quả thật không dễ dàng đối với giảng viên dạy ngôn ngữ chuyên ngành bởi vì nó không chỉ đòi hỏi về kiến thức ngôn ngữ mà còn cả về máy tính. Một vấn đề nữa đặt ra là vậy khối ngữ liệu điện tử là công cụ hay là một phương pháp nghiên cứu. Câu trả lời là điều đó phụ thuộc vào mục đích nghiên cứu của nhà ngôn ngữ học.

Nói cách khác, khối ngữ liệu định hướng giảng dạy rất hữu dụng cho giảng dạy ngôn ngữ chuyên ngành LSP. Trên thực tế, các nhà ngôn ngữ học thường sử dụng các khối ngữ liệu nhỏ dựa trên các nhu cầu đặc biệt ví dụ như khối ngữ liệu dịch vụ tài chính Hongkong, khối ngữ liệu kỹ thuật Hongkong,... Chính vì vậy, tại Trường ĐHCNH, nơi có nhiều chuyên ngành đào tạo kỹ thuật khác nhau như cơ khí, điện tử, ô tô,... thì rất cần có các khối ngữ liệu chuyên ngành bằng tiếng Anh. Tuy nhiên, để làm được điều đó cần sự phối hợp giữa các chuyên ngành và đòi hỏi nhiều công sức.

Từ thực tiễn các trường giảng dạy ngoại ngữ, chúng tôi thấy rằng sinh viên năm cuối học chuyên ngành dịch thường học ngoại ngữ dựa trên từ mới, cấu trúc từ vựng, liên kết từ,... Đối với các em thì Google là một công cụ không thể thiếu. Nhưng chúng tôi cho rằng đó không phải là một kho ngữ liệu hoàn hảo. Nếu các em muốn tìm hiểu từ trong bối cảnh thực thì phải tạo ra được khối ngữ liệu riêng và các chỉ mục ngữ cảnh (*concordance*). Thực tế là các em có thể tạo ra các khối ngữ liệu người học với quy mô 300.000 - 500.000 từ tùy vào mục đích học tập của mình. Hoặc, các em có thể sử dụng các khối ngữ liệu lớn như BNC, Lancaster Corpus, Le Monde Corpus, v.v.

Khối ngữ liệu có ý nghĩa quan trọng trong dịch máy cũng như trong nghiên cứu và thực hành dịch. Chúng cung cấp cho các dịch giả công nghệ, công cụ và dữ liệu hỗ trợ cho quá trình dịch. Vì vậy, xây dựng được một kho ngữ liệu đáng tin cậy cho dịch thuật là vô cùng cần thiết. Để làm được điều đó



Ảnh 2. Trường hợp dịch từ “thanks to” từ tiếng Anh sang tiếng Pháp (Nguồn: [tradoit.com](http://tradoit.com))

so sánh 200 triệu từ thì đây là một công cụ vô cùng



Ảnh 3. Khối ngữ liệu dịch Pháp - Anh đối với từ “grâce à” (Nhờ vào) (Nguồn: [tradoit.com](http://tradoit.com))

thì văn bản cung cấp dù dưới dạng điện tử hay in ấn đều cần chuẩn mực và đáng tin cậy. Sau đó, các khối ngữ liệu có thể được làm giàu bằng cách phân tách từ, loại hình hóa từ và phân tích cú pháp. Nói một cách khác thì khối ngữ liệu song song cần được giống hệt một cách chính xác.

Đây có thể coi là những giai đoạn đầu tiên của dịch thuật trên cơ sở khối ngữ liệu cả về nghiên cứu lí thuyết, thử nghiệm và phân tích các đặc điểm ngôn ngữ. Hiện nay, mục tiêu nghiên cứu rất đa dạng và có nhiều thay đổi. Các dịch giả muốn đào sâu các văn bản dịch bằng cả phương pháp dịch dựa trên khối ngữ liệu và tìm hiểu xem dịch thuật góp phần như thế nào vào sự phát triển ngôn ngữ. Tymoczko nhận xét về sự phát triển của khoa học ngữ liệu và dịch thuật dựa trên khối ngữ liệu như sau “đại diện cho một sự đầu tư dài hạn về nghiên cứu dịch thuật” [22].

## 6. Kết luận

Khối ngữ liệu là một công cụ tốt cho nghiên cứu từ vựng và dịch thuật. Nó giúp dịch giả và các nhà ngôn ngữ học phân tích nghĩa và văn cảnh mà từ đó diễn đạt. Qua đó, họ hiểu được bức tranh hiện thực của từ vựng đó trong hoàn cảnh chân thực nhất. Vì vậy, sử dụng khối ngữ liệu trong dịch thuật là một sự nghiên cứu nghiêm túc và cần ứng dụng vào các trường đào tạo ngôn ngữ như Baker từng nhận xét “rất cần thiết khám phá khả năng tiềm ẩn của việc sử dụng khối ngữ liệu trong nghiên cứu dịch thuật” [5]. Đối với sinh viên chuyên ngành ngoại ngữ và ngôn ngữ, phương pháp tiếp cận dựa trên khối ngữ liệu góp một phần vào các phương pháp truyền thống trong học ngoại ngữ và thực hành dịch.

## TÀI LIỆU THAM KHẢO

### Tiếng Việt

1. Đào Hồng Thu. *Ngôn ngữ học khối liệu và những vấn đề liên quan (quyển I)*. NXB Khoa học xã hội. 2009.
2. Đào Hồng Thu. *Ngôn ngữ học khối liệu và phương pháp tiếp cận hiện nay*. Tạp chí Từ điển học và Bách khoa thư, số 3. 2010.
3. Hoàng Văn Vân. *Nghiên cứu dịch thuật*, NXB Khoa học xã hội. 2005.

### Tiếng Anh

4. Adam K. *Corpora in English language teaching*, British Council. 2014.
5. Baker, M. *Corpora in translation studies: an overview and suggestions for future research*, Nxb Scientific Research Publishing. 1993.
6. Beeby and colleagues. *Corpus use and translating: corpus use for learning to translate and learning corpus use to translate*. John Benjamins. 2009.
7. Bowker L., Pearson J. *Working with specialized language*. Routledge. 2002.
8. Claudio F., Federico Z. *New directions in corpus-based translation studies*. Journal of Linguistics. 2015.
9. Farahani M. *New perspectives on corpus translations studies*. New voices in translations studies vol.28. 2023.
10. Gena B. *Using corpora in the language learning classroom: corpus linguistics for teachers*, Michigan. 2010.
11. Hatim B. *Communication across cultures: Translation theory and contrastive text linguistics*. University of Exeter Press. 1997.
12. Malmkjaer K. *Who can make nice a better word than pretty? Collocation, translation, and psycholinguistics*, pp213-232. Jhon Benjamins publishing company. 1993.
13. Laviosa S. *Description in the translation classroom: universals as a case in point*, pp. 119-132. Jhon Benjamins. 2008.
14. Laura Gavioli, Federico Zanettin. *Comparable corpora and translation: a pedagogic perspective*. First international conference on Corpus Use and Learning to Translate, Bertinoro. 1997.
15. Natalie Kübler. *Mettre en œuvre la linguistique de corpus à l'université, Vers une compétence utile pour l'enseignement/apprentissage des langues? Les cahiers de l'Acedle*. 2014.
16. Natalie Kübler. *Working with different corpora in translation teaching*, New trends in Corpora and Language Learning, London. 2011.
17. Natalie Kübler. *Corpora and LSP translation*, In Zanettin, F., S. Bernardini & D. Stewart (eds), *Corpora in Translator Education*. Manchester: St Jerome. 2003.
18. Schimied J. *Explicitness as a universal feature of translation*. Hildergard. 1996.
19. Sinclair J. *Corpus, concordance, collocation*. Oxford University Press. 1991.
20. Scott M. *PC analysis of keywords - And key keywords*. System, p. 233-245. 1997.
21. Stubbs. M.. *Text and corpus analysis*. Oxford Blackwell Publisher Ltd. 1996.
22. Tymoczko M. *Computerized corpora and the future of translation studies*. Meta, p. 652-660. 1998.