

KHẢO SÁT CÁC CÁCH THỨC TÌM KIẾM VÀ PHÂN TÍCH NGŨ LIỆU SONG NGŨ TRONG QUÁ TRÌNH HỌC HỌC PHẦN NGÔN NGỮ HỌC ĐỐI CHIẾU CỦA SINH VIÊN NĂM 3 KHOA NGÔN NGỮ VÀ VĂN HÓA NHẬT BẢN, TRƯỜNG ĐẠI HỌC NGOẠI NGỮ, ĐẠI HỌC HUẾ

Nguyễn Thị Hồng Hoa*

Trường Đại học Ngoại ngữ, Đại học Huế

Nhận bài: 08/06/2022; Hoàn thành phản biện: 29/07/2022; Duyệt đăng: 31/08/2022

Tóm tắt: Khi bắt đầu tiếp cận một bài tập lớn hay một đề tài nghiên cứu ngôn ngữ đối chiếu thì bước đầu tiên sinh viên cần làm đó là tìm kiếm, tập hợp và xử lý ngữ liệu song ngữ. Nghiên cứu đưa ra kết quả khảo sát việc tìm kiếm và xử lý ngữ liệu song ngữ khi học học phần Ngôn ngữ học đối chiếu Nhật-Việt của sinh năm 3 Khoa Ngôn ngữ và Văn hóa Nhật Bản, trường Đại học Ngoại ngữ, Đại học Huế. Từ đó, bài nghiên cứu đề xuất một số giải pháp nhằm giúp sinh viên tìm kiếm và phân tích ngữ liệu ứng với từng tiêu chí và từng cấp độ ngôn ngữ hiệu quả hơn.

Từ khóa: Ngữ liệu, song ngữ, phân tích, cấp độ ngôn ngữ

1. Mở đầu

Trong việc nghiên cứu, giảng dạy ngôn ngữ, ta cần thống kê, so sánh, đối chiếu để tìm ra các quy luật của ngôn ngữ, quy luật chuyển ngữ, các điểm tương đồng và dị biệt ở các bình diện khác nhau, các cấp độ khác nhau giữa các ngôn ngữ. Nhưng để thống kê, so sánh, đối chiếu như trên, ta cần phải có cứ liệu của các ngôn ngữ mà ta đang cần so sánh, ta gọi đó là “ngữ liệu” (corpus). *“Ngữ liệu ở đây được hiểu là tập hợp văn bản đơn ngữ, đa ngữ hay song ngữ (gồm các cặp văn bản đã được dịch thủ công, dịch tương ứng 1-1 về mặt ngữ nghĩa) và phù hợp với lĩnh vực, thể loại, niên đại mà ta cần nghiên cứu”* (Theo Đinh Điền, 2005)

Đối với sinh viên đang theo học học phần Ngôn ngữ học đối chiếu (Nhật – Việt) hiện đang được giảng dạy tại Trường Đại học ngoại ngữ, Đại học Huế thì việc tiếp cận ngữ liệu vẫn là một bước đầu tiên, quan trọng để tiến hành làm một bài tập đối chiếu. Hầu như các em đều có kiến thức nền về ngoại ngữ rất tốt (Sinh viên năm 3 – Trình độ trung, cao cấp Tiếng Nhật). Vì vậy, các em đã có thể đọc hiểu, chọn lọc và phân tích ngữ liệu ngoại ngữ khá tốt. Tuy nhiên, tuy là người Việt nhưng vì không nắm chắc được kiến thức về ngôn ngữ học và Việt ngữ học theo từng cấp độ như: từ vựng – ngữ nghĩa; ngữ pháp; văn bản..., cùng với việc không để ý đến tính “tương đương” của các ngữ liệu song ngữ (tương ứng theo cấp độ ngôn ngữ giữa hai ngữ liệu) nên Sinh viên gặp nhiều khó khăn trong việc tìm kiếm và xử lý ngữ liệu Tiếng Việt. Chính vì những điều trên đây, bài báo sẽ tập trung vào các cách tìm kiếm và phân tích hiệu quả ngữ liệu song ngữ (mà cụ thể là ngữ liệu Tiếng Việt và Tiếng Nhật).

* Email: nthhoa@hueuni.edu.vn

2. Cơ sở lý luận

2.1. Ngữ liệu

Trong một bài Nghiên cứu khoa học nói chung và nghiên cứu ngôn ngữ nói riêng thì không thể phủ nhận vai trò quan trọng của thao tác làm việc trên ngữ liệu. Việc tìm kiếm, thu thập và nghiên cứu ngữ liệu phục vụ các đề tài nghiên cứu khoa học phải đạt các yêu cầu sau:

- *Tầm tham khảo **đủ rộng** để bao quát phạm vi nghiên cứu;*
- *Mức độ tham khảo **đủ sâu**, tương ứng với cấp độ nghiên cứu;*
- *Thông tin có **chọn lọc** sao cho phù hợp với một đề tài khoa học.*

- *Thông tin **trương đối cập nhật** để đánh giá vấn đề khách quan, kịp thời, không bị lạc hậu với dòng thông tin chuyên ngành (Trung tâm Thông tin Phát triển Việt Nam (2010), “ Kỹ năng trích dẫn và lập danh mục tài liệu tham khảo”)*

Xét riêng về ngữ liệu song ngữ và ngữ liệu đa ngữ được dùng để đối chiếu các ngôn ngữ thì có những cơ sở lý thuyết như sau:

Theo từ điển Việt – Việt (2010) thì : “Ngữ liệu là tư liệu ngôn ngữ được dùng làm căn cứ để nghiên cứu ngôn ngữ”

Theo Hoàng Như Quỳnh (2011), Nghiên cứu xây dựng kho ngữ liệu song ngữ phục vụ xử lý tiếng Việt, Luận văn chuyên ngành kỹ thuật, Trường Đại học Đà Nẵng thì:

“Ngữ liệu là những dữ liệu, cứ liệu của ngôn ngữ, tức là những chứng cứ thực tế sử dụng ngôn ngữ. Những chứng cứ sử dụng ngôn ngữ này có thể là của ngôn ngữ nói mà cũng có thể là ngôn ngữ viết. Trong đó ngữ liệu tồn tại dưới dạng ngôn ngữ viết bao gồm nhiều hình thức khác nhau như: dạng giấy, dạng điện tử.

Ngữ liệu chỉ gồm các văn bản của một ngôn ngữ gọi là ngữ liệu đơn ngữ và ngữ liệu của nhiều ngôn ngữ gọi là ngữ liệu đa ngữ. Nghiên cứu này tập trung đề cập đến ngữ liệu đa ngữ. Trong ngữ liệu đa ngữ, nếu ngữ liệu của ngôn ngữ này là bản dịch của ngôn ngữ kia gọi là ngữ liệu song ngữ” (tr.3).

Ngữ liệu song ngữ thô (chưa qua xử lý) có thể được xây dựng bằng 3 cách chính: (1) Thu thập tự động từ các website song ngữ (2) Thu thập từ các ấn phẩm song ngữ (dạng điện tử) (3) Dịch thủ công, dịch song song 1-1 theo hướng dẫn (guideline) từ các văn bản nguồn có chất lượng và đúng lĩnh vực, niên đại.

Trong ngữ liệu song ngữ, các bản dịch tương ứng của mỗi ngôn ngữ phải được đặt song song với nhau hay còn được gọi là *giống hàng với nhau* (alignment). Mức độ giống hàng này có thể ở *cấp độ văn bản* (text alignment), nghĩa là từng văn bản trong ngôn ngữ nguồn được giống (liên kết) với văn bản dịch tương ứng trong ngôn ngữ đích. Tương tự cho *cấp độ đoạn* (paragraph alignment), *cấp độ câu* (sentence alignment), *cấp độ ngữ* (phrase alignment) và sâu nhất là *cấp độ từ* (word alignment). “Ngữ liệu song song có thể gồm hai ngôn ngữ hoặc nhiều hơn hai ngôn ngữ”. Cụ thể trong nghiên cứu này là hai ngôn ngữ Nhật – Việt.

Với từng cấp độ giống hàng ở ngữ liệu song ngữ trên đây thì với đối tượng là sinh viên năm thứ 3 Khoa Ngôn ngữ và Văn hóa Nhật Bản trường Đại học Ngoại ngữ, Đại học Huế, chúng tôi tập trung vào mức độ giống hàng ở cấp độ văn bản hay còn gọi là đối chiếu bản dịch.

Tuy nhiên trong các dạng bài tập thực hành và đối chiếu ngôn ngữ thì ngữ liệu song song chỉ là một phần trong các dạng ngữ liệu đa ngữ mà giảng viên và sinh viên cần phải thu thập và tiến hành phân tích xử lý. Còn có một dạng ngữ liệu “tương đối” tức là ngữ liệu Nhật – Việt “tương đương” theo các cấp độ ngôn ngữ ứng với từng yêu cầu của đề tài.

Ngữ liệu “tương đương” còn có thể mở rộng là “tương đương” về độ khó của văn bản. Chẳng hạn: Khi đối chiếu một đoạn trích có nhiều câu phức thì sinh viên phải tập trung ở các đoạn trích ở trình độ tiếng Nhật trung, cao cấp (N1, N2) chứ không thể lấy ngữ liệu có độ chênh về kiến thức như văn bản trong quyển “ Tiếng Nhật cho mọi người” – Trình độ sơ cấp và ở ngữ liệu Tiếng Việt cũng là những đoạn trích trong văn học, pháp luật, khoa học vì có thể tập trung nhiều dạng câu khác nhau.

Ví dụ: Khi có một bài tập: “ Đối chiếu câu đơn Nhật – Việt theo mục đích giao tiếp” thì một ngữ liệu gọi là “tương đương” về kiến thức phải đáp ứng các tiêu chí sau:

Tiêu chí	Ngữ liệu tiếng Nhật	Ngữ liệu tiếng Việt
Dạng và cấp độ	Chủ yếu là hội thoại nhưng cấp độ trung cao cấp	Hội thoại trích đoạn trong văn học
Yêu cầu của đề tài	Đầy đủ một số dạng câu theo mục đích giao tiếp như: hỏi, cầu khiến, cảm thán...	Đầy đủ một số dạng câu theo mục đích giao tiếp như: hỏi, cầu khiến, cảm thán...

Dưới đây là hai ngữ liệu đáp ứng tiêu chí “tương đương” kể trên vì đáp ứng được tiêu chí về độ khó của văn bản (Hội thoại trung - cao cấp N2) và tác phẩm văn học Việt Nam và cả hai ngữ liệu đều có đầy đủ các dạng câu theo mục đích nói

Ngữ liệu tiếng Nhật	Ngữ liệu tiếng Việt
学生3人が、フルでかいけいがかり 女1: 新しい会計係、決めなきゃいけないね。 女2: そうね、どうやって決める? 男: やりたい人にやってもらえばいいんじゃない? 女1: やりたい人っていったって、そんな人いるかなあ。大変な仕事だし 女2: そうよね、くじ引きにする? 男: えー! それで、もしおれになったら困るよー。 女2: そうよねー、上田君には任せられないよねー。 女1: 前に、名簿順でやろうって決めたよね。でも、伊藤さんになってから、ずっと彼女がやると言ってくれて...。やっぱり、名簿順っていうのがいいんじゃない? 女2: そうね...それが一番公平かも。 女1: でしょ。「あのさー、伊藤さんの次って、おれなんだけど...。	Chí Phèo đầy hờ ? Lê bè vừa thôi chứ! Rồi ném bẹt năm hào xuống đất, cụ bảo hần: - Cầm lấy mà cút đi cho rảnh. Rồi làm mà ăn chứ cứ báo người ta mãi à? Hần trợn mắt chỉ tay vào mặt cụ: - Tao không đến đây xin năm hào. Thấy hần toan làm dữ cụ đành dịu giọng: - Thôi, cầm lấy vậy, tôi không còn hơn. Hần vênh cái mặt lên, rất kiêu ngạo: - Không. Tao đã báo là tao không đòi tiền. -Giỏi! Hôm nay mới thấy anh không đòi tiền. Thế anh cần gì? Hần đồng dặc: - Tao muốn làm người lương thiện! Bá Kiến cười ha hả: - Ôi tưởng gì! Tôi chỉ cần anh lương thiện cho thiên hạ nhờ. Hần lắc đầu: - Không được! Ai cho tao lương thiện? Làm thế nào cho mất được những vết mảnh chai trên mặt này? Tao không thể là người lương thiện nữa. Biết

女2: ハハハハ。がんばれ! [「日本語能力試験」対策 日本語総まとめ N2(聴解)]	không? Chỉ có một cách... biết không! Chỉ có một cách là... cái này biết không? (Trích Chí Phèo – Nam Cao)
---	---

Phân tích ngữ liệu dựa theo tiêu chí về ngữ liệu “tương đương” kể trên

Tiêu chí	Ngữ liệu tiếng Nhật	Ngữ liệu tiếng Việt
Dạng và cấp độ	Hội thoại trung cao cấp trích trong: Luyện nghe năng lực Nhật ngữ (N2)	Hội thoại trích đoạn trong văn học – Tác phẩm “Chí Phèo” – Nam Cao
Yêu cầu của đề tài	<p>Đầy đủ một số dạng câu theo mục đích giao tiếp như: hỏi, cầu khiến, cảm thán...</p> <p>学生3人が、フルでかいけいがかかり</p> <p>女1: 新しい、決めなきゃいけないね = 決めなければいけない (<i>cảm thán – ý chí</i>)</p> <p>女2: そうね、どうやって決める? (<i>câu hỏi cảm thán</i>)</p> <p>男: やりたい人にやってもらえばいいんじゃない? (<i>câu hỏi gợi ý – cảm thán</i>)</p> <p>女1: やりたい人っていったって、そんな人いるかなあ。大変な仕事だし</p> <p>女2: そうよね、くじきにする? (<i>Hỏi – gợi ý – cảm thán</i>)</p> <p>男: えー! それで、もしおれになったら困るよー。 (<i>giả định – cảm thán</i>)</p> <p>女2: そうよねー、君にはせられないよねー。 (<i>Phủ định – cảm thán - đưa ý kiến</i>)</p> <p>女1: 前に、でやろうって決めたよね。でも、</p>	<p>Đầy đủ một số dạng câu theo mục đích giao tiếp như: hỏi, cầu khiến, cảm thán...</p> <p>Chí Phèo đầy hờ? (<i>hỏi - xác nhận thông tin</i>)</p> <p>Lè bè vừa thôi chứ! (<i>cảm thán – phàn nàn</i>)</p> <p>Rồi ném bẹt năm hào xuống đất, cụ bảo hấn:</p> <p>- Cầm lấy mà cút đi cho rảnh. Rồi làm mà ăn chứ cứ báo người ta mãi à? (<i>câu hỏi tu từ</i>)</p> <p>Hấn trợn mắt chỉ tay vào mặt cụ:</p> <p>- Tao không đến đây xin năm hào. (<i>khẳng định- cảm thán</i>)</p> <p>Thấy hấn toan làm dữ cụ đành dụ giọng:</p> <p>- Thôi, cầm lấy vậy, tôi không còn hơn.</p> <p>Hấn vênh cái mặt lên, rất kiêu ngạo:</p> <p>- Không. Tao đã bảo là tao không đòi tiền</p>
	<p>さんになってから、ずっと彼女がやるって言ってくれて...。やっぱり、名簿順っていうのがいいんじゃない?</p> <p>女2: そうね...それが一番かも。 (<i>Nhận định – cảm thán</i>)</p> <p>女1: でしょ。「あのさー、伊藤さんの次って、おれなんだけど...。」</p> <p>女2: ハハハハ。がんばれ! (<i>Mệnh lệnh – động viên</i>)</p> <p>「日本語能力試験」対策 日本語総まとめ N2(聴解)</p>	<p>(<i>khẳng định- cảm thán</i>)</p> <p>- Giới! (<i>cảm thán – nhận định, mỉa mai</i>). Hôm nay mới thấy anh không đòi tiền. Thế anh cần gì? (<i>hỏi – lấy thông tin</i>)</p> <p>Hấn đồng dục:</p> <p>- Tao muốn làm người lương thiện! (<i>cảm thán – bày tỏ</i>)</p> <p>Bá Kiến cười ha hả:</p> <p>- Ôi tưởng gì! Tôi chỉ cần anh lương thiện cho thiên hạ nhờ.</p> <p>Hấn lắc đầu: (<i>Trần thuật</i>)</p> <p>- Không được! (<i>Phủ định – cảm thán</i>). Ai cho tao lương thiện? Làm thế nào cho mất được những vết mảnh chai trên mặt này? (<i>Hỏi tu từ</i>) Tao không thể là người lương thiện nữa. (<i>Khả năng – cảm thán</i>). Biết không? (<i>Hỏi tu từ</i>) Chỉ có một cách... biết không! (<i>Hỏi tu từ</i>) Chỉ có một cách là... cái này biết không? (<i>Hỏi tu từ</i>)</p> <p>(Trích Chí Phèo – Nam Cao)</p>

2.2. Các nguồn thu thập ngữ liệu song ngữ

Trong quá trình truy tìm ngữ liệu, người nghiên cứu cần khai thác tối đa những nguồn hỗ trợ truy cập thông tin, tài liệu như: thầy cô, các chuyên gia, bạn bè, các thư viện, các đơn vị chuyên ngành, các tổ chức quốc tế, các diễn đàn chuyên môn và các nguồn đáng tin cậy trên mạng,... và quan trọng là phải định hướng ngữ liệu bám sát chủ đề nghiên cứu.

Có thể tiếp cận ngữ liệu từ những nguồn cụ thể sau:

Thư viện

Nhiều thư viện ở Việt Nam hiện nay (Thư viện quốc gia, Thư viện chuyên ngành khoa học, Thư viện đại học,...) chưa có đủ một lượng tài liệu mới dồi dào, phong phú, đáp ứng nhu cầu của đông đảo các nhà nghiên cứu thuộc nhiều lĩnh vực, tuy vậy không thể bỏ qua lượng tài liệu tuy cũ nhưng có tính chất kinh điển, căn bản, đã được chọn lọc và tích lũy trong thời gian dài. Và vẫn có một xác suất không nhỏ có thể tìm thấy những tài liệu thực sự quan trọng cho một đề tài nghiên cứu.

Các loại tài liệu lưu trữ ở thư viện bao gồm các tài liệu nguyên cấp như: sách, tạp chí, tài liệu nghe nhìn, các văn bản nhà nước, các luận văn, luận án, v.v., được sắp xếp và phân loại một cách khoa học, trật tự.

Các trung tâm tài liệu

Bên cạnh hệ thống thư viện được tổ chức quy củ, chặt chẽ, các trung tâm tài liệu (của các đơn vị nghiên cứu, các tổ chức chuyên môn,...) có quy mô nhỏ hơn, nhưng bù lại, các tài liệu lưu trữ có tính đặc thù chuyên môn cao, nhất là các tài liệu tập trung về một số chủ đề chuyên biệt. Thông tin về các trung tâm này có thể tìm thấy trong danh bạ các đơn vị chuyên ngành, hoặc đôi khi có thể trên Internet.

Các tủ sách chuyên ngành: Đây là dạng "trung tâm tài liệu thu nhỏ", thường gặp ở các bộ môn hoặc khoa ở trường đại học, các phòng thí nghiệm, v.v.

Các cơ sở dữ liệu tóm tắt: Ở dưới dạng CD-ROM, thường chỉ tìm thấy bài tóm tắt, các thông tin được sắp xếp theo một cấu trúc, trật tự rõ ràng, giúp tìm kiếm dễ dàng các tài liệu nguyên cấp (phổ biến nhất là các bài báo đăng trên các tạp chí chuyên ngành). Hiện nay đa số các cơ sở dữ liệu này đều có phiên bản trực tuyến.

Các danh bạ mạng và bộ máy tìm kiếm trên Internet: cho phép tìm kiếm gần như đủ loại tài liệu được đăng trên Mạng toàn cầu (World Wide Web). Tuy nhiên, độ tin cậy của thông tin tìm được bằng các công cụ này không phải lúc nào cũng tốt, thường đòi hỏi người tìm kiếm phải biết đánh giá các kết quả tìm thấy và chọn lọc những tài liệu có giá trị liên quan đến đề tài nghiên cứu

Với sự ra đời của máy tính điện tử và nhất là trong môi trường kết nối Internet toàn cầu như hiện nay, việc tập hợp ngữ liệu song ngữ đã được tự động hoá rất nhiều. Do việc xây dựng kho ngữ liệu thủ công quá tốn kém, các nhà nghiên cứu bắt đầu tìm kiếm các phương pháp xây dựng tự động. Trên thế giới, người ta đã xây dựng được nhiều kho ngữ liệu song ngữ, như: Anh-Pháp, Anh-Hoa, v.v... Trong công trình nghiên cứu của tác giả Đinh Điền (2022), nhóm tác giả đã sử dụng kho ngữ liệu song ngữ Anh-Việt điện tử 5 triệu từ được thu thập từ các tài liệu song

ngữ thuộc lĩnh vực khoa học tự nhiên chủ yếu là tin học, điện tử viễn thông, y học v.v. Chính vì vậy, Internet trở thành một nguồn cung cấp tài liệu song ngữ rất tiềm năng và hiệu quả.

3. Phương pháp nghiên cứu

Nghiên cứu này chủ yếu dựa trên phương pháp chính là nghiên cứu định lượng và được tiến hành theo những phương pháp sau:

Phương pháp điều tra: Điều tra bằng bảng hỏi bằng biểu mẫu Google đối với sinh viên năm thứ 3 (tiến hành khảo sát trên 120 sinh viên năm thứ 3 hiện đang theo học học phần Ngôn ngữ học đối chiếu) để khảo sát về cách thức tìm kiếm và xử lý ngữ liệu song ngữ mà cụ thể là hai ngôn ngữ Nhật – Việt.

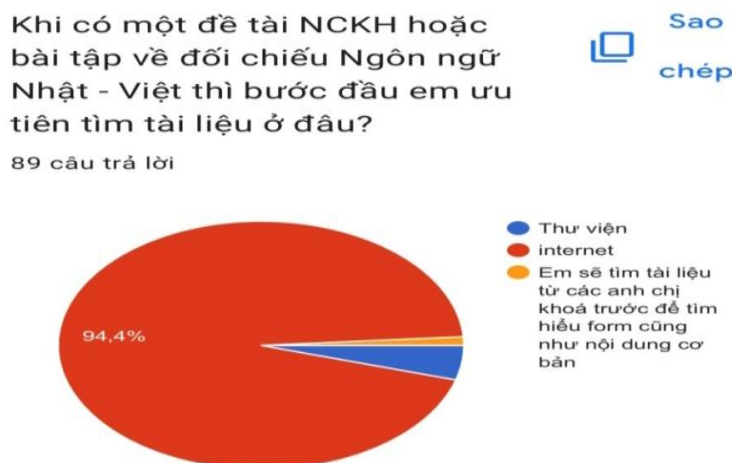
Nội dung của bảng hỏi hướng về:

- Cách thức tìm ngữ liệu để làm bài tập đối chiếu.
- Cách thức xử lý ngữ liệu của Sinh viên : (Dựa trên các khối ngữ liệu song ngữ mà các em tìm kiếm theo các cấp độ và chủ yếu là cấp độ văn bản dựa theo các tiêu chí về độ khó của văn bản song ngữ và bám sát theo yêu cầu của đề tài)
- Cách xác định độ “tương đương” của ngữ liệu và những khó khăn trong khi thu thập và xử lý ngữ liệu của sinh viên.

Phương pháp miêu tả: miêu tả kết quả khảo sát qua các sơ đồ, từ đó đề xuất cách tìm kiếm và xử lý ngữ liệu song ngữ hiệu quả đối với môn học.

4. Kết quả khảo sát

Theo kết quả khảo sát tiến hành trên hơn 120 sinh viên năm thứ 3 Khoa Ngôn ngữ và Văn hóa Nhật Bản về nơi các em ưu tiên tiếp cận ngữ liệu (Thư viện hay hay internet...) thì đại đa số (94,4%) các em ưu tiên đầu tiên đó là các nguồn tài liệu trên mạng và các phương tiện truyền thông.



Biểu đồ 1. Nơi tìm ngữ liệu ưu tiên của Sinh viên

Qua kết quả khảo sát với đại đa số sinh viên lựa chọn nơi tiếp cận đầu tiên là ngữ liệu mạng vì có thể tiếp cận nhanh hơn (chỉ cần vài thao tác đơn giản trên các công cụ nổi tiếng như Google, Ask.com). Tuy nhiên nhược điểm là có những ngữ liệu chưa được kiểm chứng bởi các chuyên gia đứng chuyên ngành hoặc quá dàn trải không thể bao quát được nội dung của một bài nghiên cứu. Điều đó gây những khó khăn trong khâu chọn lọc ngữ liệu.

Vì vậy, cần khẳng định lại rằng, nếu biết cách khai thác nghiêm túc các nguồn tài liệu truyền thống kể trên, người nghiên cứu hoàn toàn có thể tìm được những ngữ liệu có giá trị cho đề tài của mình. Vấn đề then chốt là xác định được loại tài liệu nào cần, có ở đâu, để tiếp cận được một cách hiệu quả.

Chẳng hạn như sinh viên có thể tiếp cận nguồn ngữ liệu truyền thống ngay trong các giáo trình của mình học (Đọc hiểu, nghe hiểu, ngữ pháp) và kho sách trên tủ sách chuyên ngành của Khoa và Trường.

Tuy nhiên, bên cạnh Kho ngữ liệu về ngoại ngữ rất phong phú của thư viện Trường và Khoa thì vấn đề sử dụng ngoại ngữ và ngôn ngữ để phân tích ngữ liệu song ngữ cũng là vấn đề đáng quan tâm. Bởi vậy, việc tìm ngữ liệu và thao tác để xử lý ngữ liệu phải được chú trọng.



Biểu đồ 2. Khó khăn trong vấn đề tiếp cận và phân tích ngữ liệu

Một khó khăn khi tiếp cận và phân tích ngữ liệu theo biểu đồ 2 đó là phân tích ngữ liệu liên quan đến ngoại ngữ. Điều này có thể nhận thấy những mặt hạn chế trong kỹ năng đọc hiểu ngoại ngữ của sinh viên. Có đến 66,3% sinh viên cảm thấy ngoại ngữ chưa đủ tốt để đọc và phân tích ngữ liệu.

Sau khi tìm được ngữ liệu thì bước tiếp theo em ưu tiên làm là gì?

89 câu trả lời

Phân chia các ngữ liệu theo nội dung phù hợp
Xác nhận dữ liệu có đủ điều kiện chưa
Lựa chọn những ngữ liệu chính và quan trọng
vẽ bố cục cho bài nghiên cứu
Sau khi tìm ngữ liệu thì em sẽ tổng hợp lại, đọc kĩ lại để chắt lọc những thông tin cần thiết, sát đề tài nghiên cứu ạ
Sắp xếp ngữ liệu
Chọn lọc tài liệu phù hợp, phân loại, phân tích, mổ xẻ tài liệu ngữ liệu
Lọc dữ liệu

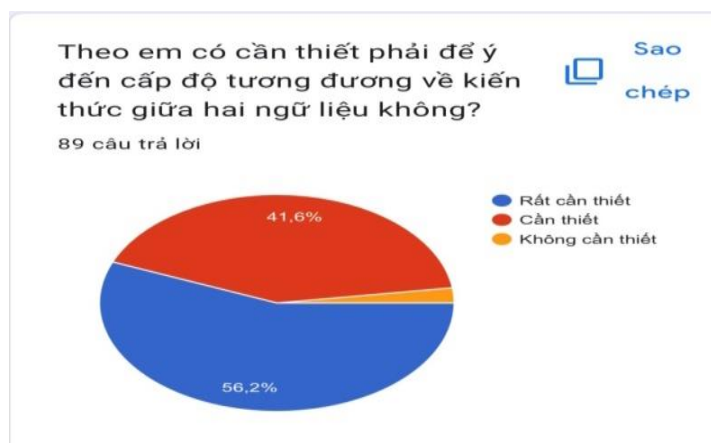
Ảnh. Cách thức xử lý ngữ liệu của sinh viên

Về cách thức xử lý ngữ liệu của sinh viên, sau khi trả lời câu hỏi mở thì đại đa số các câu trả lời đều tập trung vào các điểm sau:

- + Đọc và phân tích sắp xếp ý thành nhóm theo yêu cầu của đề tài (57%)
- + Phân chia các ngữ liệu theo nội dung phù hợp (33%)
- + Liệt kê, tập hợp và sắp xếp ngữ liệu theo yêu cầu cụ thể của đề tài (43,5%)

Qua kết quả khảo sát: Đại đa số sinh viên đều biết cách đọc và phân tích ngữ liệu. Tuy nhiên, kỹ năng phân tích cụ thể ngữ liệu về mặt ngoại ngữ, việt ngữ học và các cấp độ ngôn ngữ đang còn hạn chế.

Về mức độ “tương đương” về kiến thức giữa hai ngữ liệu để đối chiếu thì đại đa số Sinh viên được khảo sát cho là rất cần thiết (56,2%) và cần thiết 41,6%). Qua đó thể hiện các em đã hiểu cách tìm một ngữ liệu tương ứng cả về cấp độ ngôn ngữ, độ khó, độ dài và chi tiết, đầy đủ các yêu cầu của từng đề tài thể hiện trên ngữ liệu để việc làm việc trên ngữ liệu được hiệu quả hơn.



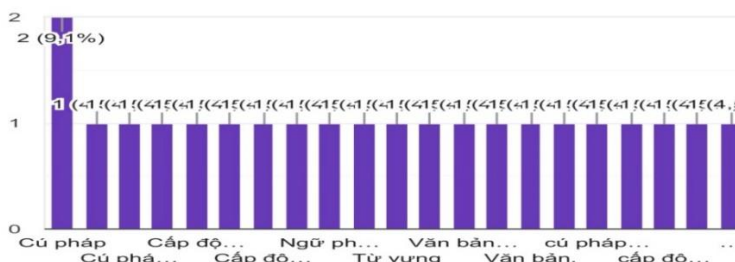
Biểu đồ 3. Về cấp độ tương đương về kiến thức của ngữ liệu

Về việc khó khăn khi tìm kiếm ngữ liệu ứng với từng cấp độ ngôn ngữ thì đại đa số sinh viên đều cảm thấy khó khăn ở cấp độ cú pháp. Điều đó có thể lý giải rằng ở cấp độ này thì kỹ năng phân tích câu và các thành phần câu của Sinh viên đang còn hạn chế.

Trong ba cấp độ ngôn ngữ (Từ vựng, ngữ nghĩa; Cú pháp; văn bản) thì em cảm thấy khó khăn khi tìm kiếm và phân tích ngữ liệu ở cấp độ nào? Vì sao?



22 câu trả lời



Biểu đồ 4. Khó khăn trong việc tìm kiếm ngữ liệu ứng với từng cấp độ của ngôn ngữ

5. Thảo luận và đề xuất

5.1. Đối với việc ưu tiên tìm kiếm ngữ liệu truyền thống.

Thay vì internet và các công cụ mạng, giảng viên nên định hướng cho sinh viên những ưu điểm về sự chính xác và tin cậy của ngữ liệu truyền thống thông qua các học phần chuyên ngành liên quan đến Nghiên cứu khoa học, ngôn ngữ đối chiếu nhằm giúp sinh viên có thêm kiến thức và định hướng rõ ràng hơn trong bước đầu tiên của một đề tài nghiên cứu.

5.2. Đối với kỹ năng tập hợp, liệt kê và phân tích ngữ liệu

Trong quá trình giảng dạy thì giảng viên cần cho sinh viên thực hành trên nhiều các dạng bài tập trên nhiều cấp độ của ngôn ngữ để sinh viên có thể rèn luyện các kỹ năng trên một cách hiệu quả.

Nên tập trung thực hành nhiều vào kỹ năng đọc hiểu kể cả ngoại ngữ và Việt ngữ học làm cơ sở cho việc phân tích tốt ngữ liệu của sinh viên

Giảng viên nên chú trọng vào các cấp độ ngôn ngữ mà sinh viên còn hạn chế trong khi phân tích ngữ liệu chẳng hạn như cấp độ cú pháp dựa vào kết quả khảo sát được nêu trên đây.

5.3. Đề xuất các thao tác phân tích ngữ liệu song ngữ Nhật – Việt ứng với từng cấp độ của ngôn ngữ

Vì các dạng bài tập về ngôn ngữ đối chiếu phủ đều trên các cấp độ ngôn ngữ nên Sinh viên phải có kỹ năng làm việc trên ngữ liệu theo từng cấp độ. Dựa vào kết quả khảo sát ở biểu đồ 4 có thể phân loại ra từng cấp độ ngôn ngữ cụ thể như: từ vựng – ngữ nghĩa; cú pháp – ngữ dụng; văn bản. Có thể thấy sinh viên đại đa số thấy khó khăn trong việc phân tích ngữ liệu ở cấp độ ngữ pháp - ngữ dụng, tuy nhiên những đề xuất về các thao tác phân tích ngữ liệu theo các cấp độ rất cần thiết để giúp Sinh viên hoàn thiện kỹ năng làm việc trên ngữ liệu được tốt hơn.

5.3.1. Các thao tác phân tích ngữ liệu song ngữ Nhật – Việt theo cấp độ từ vựng – ngữ nghĩa

* Yêu cầu về kiến thức nền: Trước hết sinh viên phải có được kiến thức nền về ngôn ngữ liên quan đến mảng từ vựng - ngữ nghĩa Nhật – Việt.

(Chẳng hạn như hệ thống từ loại: Danh – Động – Tính, cấu tạo từ, chuyên loại từ, cụm từ, ...)

Ví dụ: Với đề tài: “Đổi chiều các hiện tượng chuyển loại từ Nhật – Việt” có thể tìm ngữ liệu đoạn trích như có chứa những hiện tượng chuyển loại từ (như : động từ → danh từ; tính từ → danh từ)

Chẳng hạn như hai cột ngữ liệu sau sẽ đáp ứng được yêu cầu của đề tài

Tiêu chí	Ngữ liệu tiếng Nhật	Ngữ liệu tiếng Việt
Dạng và cấp độ	Ngữ liệu đoạn trích văn bản trung cao cấp trích trong: Luyện đọc năng lực Nhật ngữ (N3)	Ngữ liệu đoạn trích trong văn bản báo chí
Yêu cầu của đề tài	Tập hợp được các dạng chuyển đổi từ loại :(Động từ → danh từ; tính từ → danh từ...) 人間とは不思議なもの(tính từ →danh từ)で、年をとってもいろいろな 新しいこと (tính từ →danh từ)を経験したいという気持ちがある。例えば自分で茶わんを作ってみるとか、ピザを焼いてみるとか、さまざまな 新しい がある。しかし、当然のことながら、それがうまくいくこと(động từ →danh từ)は少ない。茶わんはぜんぜん丸く作れないし、ピザはおいしくない。だが、それを恥ずかしく思う必要はないのだ。最初からうまくできるなら、世の中のさまざまな学校や修業は必要ないことになる。初めはうまくいかないの(tính từ → động từ → danh từ)が当たり前なのだ。それなのに、家族の者に「変な形」、「まずい」などと言われると、「もう二度と作らない」と思ってしまう。初めての 挑戦 (tính từ →danh từ)には、周りの人々の理解と応援も必要なようだ。	Tập hợp được các dạng chuyển đổi từ loại :(Động từ → danh từ; tính từ → danh từ...) Khi tăng sự phát triển (động từ → danh từ) kinh tế và lợi ích xã hội điều đó đồng nghĩa với việc gia tăng (động từ → danh từ) khoảng cách cho 26 triệu trẻ em Việt Nam trong việc tiếp cận (động từ → danh từ) các dịch vụ và hỗ trợ, với những tác động (động từ → danh từ) nghiêm trọng đến sức khỏe và dinh dưỡng của trẻ em. Mặc dù đã có những tiên bộ (tính từ → danh từ) đáng kể thể hiện trên một loạt các chỉ số về sức khỏe trong hai thập kỷ qua, sự bất bình đẳng (tính từ → danh từ) và khác biệt vẫn còn ở khu vực dân tộc thiểu số, thu nhập hộ gia đình, những khó khăn (tính từ → danh từ) của cuộc sống ở nông thôn và miền núi và giáo dục cho người mẹ. <i>Trích từ: Bài báo “Vi sự sống còn và phát triển của trẻ em”</i>
	[日本語能力試験] 新完全マスターN3 (読解)	

5.3.2. Các thao tác phân tích ngữ liệu song ngữ Nhật – Việt theo cấp độ cú pháp – ngữ dụng

Cấp độ cú pháp – ngữ dụng cũng là một trong các phạm trù phổ biến mà giáo viên có thể giao các bài tập thực hành và gợi ý các đề tài tiểu luận cho sinh viên.

* Yêu cầu về kiến thức nền:

Người học phải nắm vững các hình thái câu trong tiếng Nhật và tiếng Việt: Các dạng câu đơn, ghép, phức; và các phạm trù từ vựng – ngữ pháp xoay quanh các kiểu câu như: câu theo mục đích giao tiếp (trần thuật, nghi vấn, cảm thán, mệnh lệnh v.v)

Chẳng hạn :

Về cấu trúc câu thì trật tự từ của câu tiếng Nhật sẽ là :

S – O – V (động từ luôn đứng sau cùng và gắn liền với bổ ngữ sẽ là các trợ từ, chủ thể phần lớn có thể lược bỏ)

+ (私は) 毎日ご飯を作っています (Tôi nấu ăn hàng ngày)

S O V

Còn trật tự từ ở ngữ pháp truyền thống trong tiếng Việt sẽ là:

S – V – O

+ Sáng nay, tôi đi học

S V O

Ví dụ: Với đề tài: “Đối chiếu cách đặt tiêu đề quảng cáo trong Tiếng Nhật và tiếng Việt dưới góc độ câu theo mục đích giao tiếp” thì đầu tư vào việc tìm kiếm và xử lý ngữ liệu cần nhiều thời gian và phải có kỹ năng tập hợp, sắp xếp ngữ liệu theo từng phần, mục của yêu cầu đề tài hay còn gọi là: “ngữ liệu rời” hay “ngữ liệu tổng hợp”

Tiêu chí ngữ liệu ở đề tài trên

Tiêu chí	Ngữ liệu tiếng Nhật	Ngữ liệu tiếng Việt
Dạng và cấp độ	Dạng ngữ liệu rời (tổng hợp) trong các tờ rơi quảng cáo hoặc tranh ảnh, tạp chí, báo chí Cấp độ Cụm từ - Câu (tiếng Nhật trung cao cấp)	Dạng ngữ liệu rời (tổng hợp) trong các tờ rơi quảng cáo hoặc tranh ảnh, tạp chí, báo chí Cấp độ Cụm từ - Câu (tiếng Việt ứng dụng)
Yêu cầu của đề tài	Ngữ liệu phải tổng hợp được các cấp độ câu như: trần thuật, nghi vấn, cảm thán, mệnh lệnh... và các hiệu quả tu từ để làm nổi bật tiêu đề	Ngữ liệu phải tổng hợp được các cấp độ câu như: trần thuật, nghi vấn, cảm thán, mệnh lệnh... và các hiệu quả tu từ để làm nổi bật tiêu đề

Dạng ngữ liệu này là dạng ngữ liệu tổng hợp nên trước hết nhà nghiên cứu phải liệt kê và tìm kiếm trên tất cả các tài liệu tham khảo có liên quan, sau đó chọn lọc, phân loại và bước cuối cùng là tiến hành đối chiếu.

Với đề tài kể trên thì những ngữ liệu rời như sau có thể đáp ứng được tiêu chí của đề tài. Tất nhiên với một đề tài như thế này việc tìm càng nhiều ngữ liệu rời bao nhiêu thì càng dễ dàng để phân tích và so sánh bấy nhiêu. Tuy nhiên trong phạm vi một ví dụ thì có thể tham khảo bằng ngữ liệu dưới đây:

Tiêu chí	Ngữ liệu tiếng Nhật	Ngữ liệu tiếng Việt
Dạng và cấp độ	Dạng ngữ liệu rời (tổng hợp) trong các tờ rơi quảng cáo hoặc tranh ảnh, tạp chí, báo chí Cấp độ Cụm từ - Câu (Tiếng Nhật trung cao cấp) - ここは日本一心の距離が違いサウアリパーク。(PL1)	Dạng ngữ liệu rời (tổng hợp) trong các tờ rơi quảng cáo hoặc tranh ảnh, tạp chí, báo chí Cấp độ Cụm từ - Câu (Tiếng Việt ứng dụng) - Quạt điện cao cấp Senko Làm mát cho cuộc sống (PL12) -Hải sản biển vàng

	<ul style="list-style-type: none"> - マンションはポリエステル。(PL2) - 耐震対策工事を行っています。(PL3) - 電車に乗って会いに行くね！(PL4) - 仕事の無駄よ、さようなら。(PL5) - 親切な人に見つけてもらってね(PL6) - がんばれ受験生！(PL7) - 学生さん、苦労するべからず！ 	<ul style="list-style-type: none"> Cung cấp hải sản và thực phẩm sạch (PL13) - EXIMBANK - Ngân hàng xuất nhập khẩu Việt Nam(PL14) - Ôi giờ ơi! Rê quá (PL15) - Leo Sảng khoái mỗi ngày! (PL16) -Kornets Cực “Quây” Cực Kết! (PL17) -Viettel Hãy nói theo cách của bạn. (PL18) -Nescafe Capuccino Chuẩn gu và phê quán
	<ul style="list-style-type: none"> (PL8) - ハイボーター？(PL9) - 池袋ならではの手みやげってあるかしら？(PL10) - 山陽電車の車内は寒すぎる！どうにかならないの？(PL11) 	<ul style="list-style-type: none"> Mood lên chờ chi! (PL19) - Sony Xperia go Bụi, nước, va đập Ngại gì thử thách? (PL20) - Mừng 1 tuổi GoViet Triệu ưu đãi thật tuyệt(PL21) -Giải lao rồi coca cola thôi(PL22)
Yêu cầu của đề tài	Ngữ liệu phải tổng hợp được các cấp độ câu như: trần thuật, nghi vấn, cảm thán, mệnh lệnh... và các hiệu quả tu từ để làm nổi bật tiêu đề	Ngữ liệu phải tổng hợp được các cấp độ câu như: trần thuật, nghi vấn, cảm thán, mệnh lệnh... và các hiệu quả tu từ để làm nổi bật tiêu đề
Trần thuật	<ul style="list-style-type: none"> - ここは日本一心の距離が違いサウアリパーク。(PL1) - マンションはポリエステル。(PL2) - 耐震対策工事を行っています。(PL3) 	<ul style="list-style-type: none"> - Quạt điện cao cấp Senko Làm mát cho cuộc sống(PL12) - Hải sản biển vàng Cung cấp hải sản và thực phẩm sạch (PL13) - EXIMBANK Ngân hàng xuất nhập khẩu Việt Nam (PL14)
Nghi vấn, câu hỏi tu từ	<ul style="list-style-type: none"> - ハイボーター？(PL9) - 池袋ならではの手みやげってあるかしら？(PL10) - 山陽電車の車内は寒すぎる！どうにかならないの？(PL11) 	<ul style="list-style-type: none"> - Nescafe Capuccino Chuẩn gu và phê quán Mood lên chờ chi?(PL19) - Sony Xperia go Bụi, nước, va đập Ngại gì thử thách? (PL20)
Cảm thán	<ul style="list-style-type: none"> - 仕事の無駄よ、さようなら。(PL5) - 親切な人に見つけてもらってね 	<ul style="list-style-type: none"> - Ôi giờ ơi! Rê quá(PL15) - Leo Sảng khoái mỗi ngày! (PL16)
	(PL6)	<ul style="list-style-type: none"> - Kornets Cực “Quây” Cực Kết! - Mừng 1 tuổi Go Viet Triệu ưu đãi thật tuyệt (PL21)

Mệnh lệnh	<ul style="list-style-type: none"> - 電車に乗って会いに行くね！ (PL4) - がんばれ受験生！ (PL7) - 学生さん、苦勞するべからず！ (PL8) 	<ul style="list-style-type: none"> - Viettel Hãy nói theo cách của bạn. (PL18) - Giải lao rồi coca cola thôi(PL22)
-----------	---	---

5.3.3. Các thao tác phân tích ngữ liệu song ngữ Nhật – Việt theo cấp độ văn bản – phong cách chức năng của văn bản

Ở cấp độ này, thường có các đề tài liên quan đến đối chiếu bản dịch hoặc đối chiếu các phong cách chức năng của văn bản như: (phong cách báo chí – công luận; phong cách hành chính – công vụ; phong cách nghệ thuật; phong cách sinh hoạt hàng ngày)

** Yêu cầu về kiến thức nền:*

Ở cấp độ văn bản, đòi hỏi người học phải nắm thật vững cấu trúc câu, các phương tiện liên kết trong văn bản. Người học cũng cần nắm vững các phong cách chức năng của ngôn ngữ, cách lựa chọn và sử dụng từ ngữ theo đúng phong cách chức năng như: báo chí, hành chính, sinh hoạt hàng ngày v.v... Biết cách đối chiếu sau khi đã liệt kê và phân tích ngữ liệu.

Ví dụ : Đối chiếu các bản dịch song ngữ (Về phong cách hành chính)

Tiêu chí cho ngữ liệu song ngữ ở đề tài trên

Tiêu chí	Ngữ liệu tiếng Nhật	Ngữ liệu tiếng Việt
Dạng và cấp độ	Dạng ngữ liệu đoạn trích trong văn bản (có thể là ngữ liệu tiếng Nhật là văn bản nguồn) Cấp độ tiếng Nhật trung cao cấp	Dạng ngữ liệu đoạn trích trong văn bản (có thể là bản dịch từ văn bản nguồn tương ứng) Các văn bản chuyên ngành sâu như: pháp luật, báo chí...
Yêu cầu của đề tài	Ngữ liệu phải tập trung vào các văn bản hành chính chuyên ngành (pháp luật, khiếu nại...) Phải có các từ cụm từ “thuật ngữ” chuyên ngành,	Ngữ liệu phải tập trung vào các văn bản hành chính chuyên ngành (pháp luật, khiếu nại...) Phải có các cụm từ “thuật ngữ” chuyên ngành, từ ghép, từ Hán Việt.
	Phải có các cấu trúc câu đặc trưng cho phong cách hành chính công vụ như: Kiểu câu khuyết chủ ngữ, câu ghép. Cách ngắt dòng, xuống dòng đặc trưng...	Phải có các cấu trúc câu đặc trưng cho phong cách hành chính công vụ như: Kiểu câu khuyết chủ ngữ, câu ghép. Cách ngắt dòng, xuống dòng đặc trưng...

Phân tích 2 ngữ liệu song ngữ về văn bản pháp luật dưới đây để thấy mức độ đáp ứng được tiêu chí của đề tài

Tiêu chí	Ngữ liệu tiếng Nhật	Ngữ liệu tiếng Việt
Dạng và cấp độ + Đoạn trích văn bản Hành chính – công vụ (Pháp luật) + Dóng hàng Việt – Nhật	の(a) び の(b) -は、であるが、められた おえてにがにしたというである。 (c)	Cơ sở tuyên án. (a) Ví dụ về việc vi phạm luật xuất nhập cảnh và công nhận người tị nạn. (b)

(Ứng với từ gạch đầu dòng sẽ có ý đóng hàng song song)	<p>-ががににしたがりのであることなどにらすと,のはい。(d)</p> <p>-で,は,についてのをし,は,にってまじめなをりながら,ちっていくことをっていること,とを にしていたが,のでとしてにするちでおり,にするなをえていることなど,にとつてむべきもある。(e)</p> <p>-そこで,これらのをしてのとおりをした。(f)</p> <p>(ハンドブック実践編 [ベトナム語] (改訂版) trang (116-117)</p>	<p>- Đây là vụ án mà bị cáo là công dân của nước Y đã lưu trú bất hợp pháp tại Nhật Bản quá thời gian cho phép. (c)</p> <p>- Số thời gian lưu trú bất hợp pháp của bị cáo tại Nhật Bản quá thời hạn cho phép là trên 2 năm nên trách nhiệm hình sự của bị cáo rất nặng. (d)</p> <p>- Mặt khác, việc bị cáo đã bày tỏ sự ân hận về việc phạm tội này và đã hứa sau khi trở về nước sẽ cố gắng làm việc, tự sửa đổi mình; vợ hôn thê (hôn phu) của bị cáo hiện đang sống chung đã có quyết tâm lập gia đình với bị cáo tại đất nước của bị cáo và mong được tòa khoan hồng cho tội trạng này. (e)</p>
		<p>- Suy xét tổng hợp những tình huống trên, tòa quyết định mức án như đã tuyên bố trong văn bản chính thức. (f)</p> <p>ハンドブック実践編 [ベトナム語] (改訂版), trang (117 -118)</p>
<u>Yêu cầu của đề tài</u>		
Phải có các từ cụm từ, “thuật ngữ” chuyên ngành		Tuyên án; luật xuất nhập cảnh; vi phạm; bị cáo; trách nhiệm hình sự...
Sử dụng nhiều từ ghép và từ ghép Hán – Việt (ở ngữ liệu tiếng Việt)		Tuyên án; vi phạm; bị cáo; trách nhiệm hình sự; Vợ hôn thê (hôn phu); quyết tâm; quyết định...
Có các cấu trúc câu đặc trưng cho phong cách hành chính công vụ như: Kiểu câu khuyết chủ ngữ, câu ghép...	Câu (b): Khuyết chủ ngữ Câu (c;e;f): Ghép Câu (d): Phức	Câu (b): Khuyết chủ ngữ Câu (c;e;f): Ghép Câu (d): Phức
Cách ngắt dòng, xuống dòng đặc trưng...	Cách ngắt dòng xuống đoạn ngắn gọn, rõ ràng, đặc trưng cho phong cách Hành chính - công vụ	Cách ngắt dòng xuống đoạn ngắn gọn, rõ ràng, đặc trưng cho phong cách Hành chính - công vụ

6. Kết luận

Nghiên cứu đưa ra kết quả khảo sát thực trạng của sinh viên năm 3 Khoa Ngôn ngữ và Văn hóa Nhật Bản, Trường Đại học Ngoại ngữ, Đại học Huế trong việc tìm kiếm và xử lý ngữ liệu song ngữ phục vụ cho việc học học phần Ngôn ngữ học đối chiếu (Nhật – Việt). Nghiên cứu cũng đã đi sâu chỉ rõ các tiêu chí đáp ứng theo từng đề tài cụ thể và các ví dụ điển hình trong việc phân

tích ngữ liệu theo từng cấp độ ngôn ngữ cụ thể như: từ vựng – ngữ nghĩa, ngữ pháp – ngữ dụng, văn bản. Nghiên cứu cũng đưa ra các thảo luận và đề xuất cụ thể trong việc nâng cao kỹ năng làm việc trên ngữ liệu của sinh viên như: trang bị kiến thức nền về ngôn ngữ, Việt ngữ học và ngoại ngữ; khuyến khích sinh viên nghiên cứu khoa học và làm các dạng bài tập đối chiếu đa dạng theo từng cấp độ của ngôn ngữ; Tập trung vào cấp độ ngôn ngữ mà sinh viên còn bị hạn chế. Nghiên cứu có thể là tài liệu tham khảo cho người dạy và người học Tiếng Nhật trong quá trình nghiên cứu, giảng dạy và học tập của bản thân.

Tài liệu tham khảo

- Bộ Giáo dục và Đào tạo (2019). *Sách giáo khoa Ngữ văn 11* (Tập 1). Nhà xuất bản Giáo dục
- Bùi Mạnh Hùng (2008). *Ngôn ngữ học đối chiếu*. Nhà xuất bản Giáo dục.
- Diệp Quang Ban (2005). *Ngữ pháp tiếng Việt*. Nhà xuất bản Giáo dục.
- Đình Điền & Lý Ngọc Minh (2015). Ứng dụng Ngữ liệu Song ngữ Anh-Việt trong Giảng dạy Ngôn ngữ. *Hội thảo Liên ngành Ngôn ngữ học Ứng dụng & Giảng dạy Ngôn ngữ* (tr.559-567). Huế.
- Đình Điền (2005). *Xây dựng và khai thác ngữ liệu song ngữ Anh-Việt điện tử*. Luận án Tiến sĩ Ngôn ngữ học so sánh. Đại học Khoa học Xã hội & Nhân văn, ĐHQG Tp. HCM.
- Hoàng Như Quỳnh (2011). *Nghiên cứu xây dựng kho ngữ liệu song ngữ phục vụ xử lý tiếng Việt*. (Luận văn Chuyên ngành Kỹ thuật. Trường Đại học Đà Nẵng)
- Luyện thi năng lực Nhật ngữ N2 (nghe hiểu) 「日本語能力試験」対策 日本語総まとめN2 (聴解).
- Luyện thi năng lực Nhật ngữ N3 (đọc hiểu) 「日本語能力試験」新完全マスターN3 (読解).
- Nhiều tác giả (2016). *Văn mới 2015-2016* (tr.298; 23cm). TP. Hồ Chí Minh: Trẻ.
- Tác giả mới & tác giả đang được mến mộ/Hồ Anh Thái tuyển (2017). *Văn Mới 2016-2017* (In lần thứ 1) (tr.288; 23cm). TP. Hồ Chí Minh: Trẻ.
- Trung tâm Thông tin Phát triển Việt Nam (2010). *Kỹ năng trích dẫn và lập danh mục tài liệu tham khảo*.
- Sasaki Hitoko & Matsumoto Noriko (Phi Ngọc dịch) (2019). *Luyện thi năng lực Nhật ngữ N1 - Đọc hiểu*. TP. Hồ Chí Minh.
- の花束 検索(T52-53).
- ハンドブック実践編 [ベトナム語] (改訂版).
- Cum từ cố định* (2010). Truy cập ngày 10/09/2021 từ: <https://vietnghuhoc.violet.vn/entry/cum-tu-co-dinh-4524629.html>.
- ハンドブック実践編 [ベトナム語] (改訂版)(2021). Truy cập ngày 10/09/2021 từ: <http://www.hosokai.or.jp/>.
- Định nghĩa ngữ liệu*. Truy cập ngày 01/09/2021 <https://vtudien.com/viet-viet/dictionary/nghia-cua-tu-ng%E1%BB%AF%20li%E1%BB%87u>.
- Slogan 50 thương hiệu nổi tiếng của Việt Nam* (2018). Truy cập ngày 12/09/2021 từ: <http://greenwaycreative.vn/slogan-50-thuong-hieu-noi-tieng-cua-viet-nam.html>.
- Vì sự sống còn và phát triển của trẻ em* (2018). Truy cập ngày 12/09/2021 từ: <https://www.unicef.org/vietnam/vi/v%C3%AC-s%E1%BB%B1-s%E1%BB%91ng-c%C3%B2n-v%C3%A0-ph%E1%BB%83n-c%E1%BB%A7a-tr%E1%BA%BB-em>.

**AN INVESTIGATION IN SEARCHING AND PROCESSING
BILINGUAL CORPUS IN PROCESS OF LEARNING
CONTRASTIVE LINGUISTICS OF THIRD – YEAR STUDENTS
OF DEPARTMENT OF JAPANESE LANGUAGE AND CULTURE,
UNIVERSITY OF FOREIGN LANGUAGES
AND INTERNATIONAL STUDIES, HUE UNIVERSITY**

Abstract: In order to begin the approach of a contrastive linguistic assignment or topic, what students need to do at the first step is to search, collect and process bilingual corpus. The study shows the results of survey on searching and processing bilingual corpus in process of learning Japanese – Vietnamese contrastive linguistics of third – year students of Department of Japanese Language and Culture, Hue University of Foreign Languages and International Studies, Hue University. From the results, the study proposes some solutions to help students to search and effectively analyse corpus corresponding to each criterion and each language level.

Keywords: Corpus, bilingual, analyse, language levels