



ỨNG DỤNG KỸ THUẬT ĐO ĐỘ KHÓ TRONG BIÊN SOẠN CÁC DẠNG BÀI TẬP KỸ NĂNG VIẾT TIẾNG VIỆT TRÌNH ĐỘ A1, A2, B1 THEO “KHUNG NĂNG LỰC TIẾNG VIỆT DÙNG CHO NGƯỜI NƯỚC NGOÀI”

¹Dương Thị Nhung, ¹Hồ Viết Hoàng, ¹Nguyễn Thị Hoài Thanh, ²Đinh Điền

¹Trường ĐH Ngoại ngữ, Đại học Huế, 57 Nguyễn Khoa Chiêm, tp. Huế, Việt Nam

²Trường ĐH Khoa học Tự nhiên, ĐHQG thành phố Hồ Chí Minh, Việt Nam

* Tác giả liên hệ: **Dương Thị Nhung** < dtnhung@hueuni.edu.vn >

(Ngày nhận bài: 02-10-2023; Ngày chấp nhận đăng: 06-11-2023)

Tóm tắt. Nghiên cứu về kỹ thuật đo độ khó có ý nghĩa thực tiễn trong việc đưa ra các giải pháp nhằm đáp ứng nhu cầu giảng dạy ngoại ngữ, trong đó có tiếng Việt cho người nước ngoài. Kết quả nghiên cứu đã phân tích một số yếu tố ảnh hưởng tới độ khó trong quá trình giảng dạy kỹ năng viết tiếng Việt dưới cấp độ từ, câu và văn bản. Ngoài ra, bài viết còn tổng hợp được 30 dạng bài tập có thể sử dụng trong đề thi viết. Kết hợp với việc vận dụng phần mềm, kỹ thuật đo độ khó văn bản của khoa học máy tính để xây dựng được 900 bài tập tương ứng với 4.500 câu hỏi nhằm phục vụ cho việc giảng dạy kỹ năng viết nói riêng và biên soạn giáo trình, đề thi đánh giá năng lực tiếng Việt cho người nước ngoài nói chung.

Từ khóa: tiếng Việt; kỹ năng Viết; độ khó; trình độ A1, A2, B1

APPLICATION OF TEXT READABILITY TECHNIQUES IN COMPILATION SOME TYPES OF VIETNAMESE WRITING EXERCISES BASE ON “VIETNAMESE COMPETENCY FRAMEWORK FOR FOREIGNERS” LEVEL A1, A2, B1

Dương Thị Nhung, Ho Viet Hoang, Nguyen Thi Hoai Thanh, Dinh Dien

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

*Correspondence to **Dương Thị Nhung** < dtnhung@hueuni.edu.vn >

(Received: October 02, 2023; Accepted: November 06, 2023)

Abstract. The study of text readability techniques play an important role in finding the solutions to meet the language teaching demand, including Vietnamese for foreigners. The paper’s result investigated and analysed some factors having influence on readability in teaching Vietnamese writing skill with 3 main levels of words, sentences and texts. In addition, the combination of text readability techniques and 30 writing test forms, which are collected and classified by author may help this paper creates 900 exercises with 4.500 questions. Its value has effectively serve for teaching Vietnamese writing skill in particular and writing textbooks, tests to assess Vietnamese language ability for foreigners in general.

Keywords: Vietnamese; writing skill; readability; A1, A2, B1 level.

1. Đặt vấn đề

Hiện nay, nhu cầu học tập và nghiên cứu về tiếng Việt của người nước ngoài đang ngày càng tăng cao. Nhiều trung tâm, viện nghiên cứu, các cơ sở đào tạo Việt Nam học và tiếng Việt đã ra đời ở trong nước và trên thế giới nhằm phục vụ cho nhu cầu nghiên cứu, học tập về văn hoá Việt Nam và tiếng Việt. Thế nhưng, dường như giữa các cơ sở đào tạo này vẫn chưa tìm thấy điểm chung trong việc thống nhất các tiêu chuẩn, tiêu chí đánh giá năng lực người học dựa

trên những kết quả của việc ứng dụng nghiên cứu ngôn ngữ học máy tính cũng như việc cấp các chứng chỉ liên quan, quỹ đề thi còn mang tính nội bộ.

Trong xu thế ấy, chúng tôi nhận thấy việc ứng dụng đo độ khó ngữ liệu nhằm áp dụng vào giảng dạy các kỹ năng thực hành tiếng Việt và xây dựng quỹ đề thi để đánh giá theo đúng tiêu chuẩn và năng lực người học là vô cùng cần thiết. Việc đánh giá này sẽ tạo thêm động lực cho cả người dạy và người học khi họ nắm được mức độ truyền tải và tiếp nhận kiến thức để điều chỉnh sao cho phù hợp.

“Độ khó của văn bản là tập hợp tất cả các yếu tố nội tại của văn bản khiến cho văn bản dễ hay khó tiếp thu bởi đa số người đọc ở một cấp độ cụ thể. Độ khó của văn bản phụ thuộc vào độ phổ biến của từ vựng (tuần suất), độ phức tạp của cấu trúc câu (số lượng mệnh đề, chiều dài câu) và tổ chức văn bản (liên kết văn bản)” [2, Tr.263]. Một văn bản càng dễ đọc và nội dung rõ ràng thì càng có nhiều khả năng thu hút và giữ được sự hứng thú của người đọc. Ngược lại, văn bản khó đọc làm người đọc mất đi tính kiên nhẫn và động lực để tiếp tục. Đã có nhiều nghiên cứu đạt thành tựu về đánh giá độ khó của văn bản trong tiếng Anh và một số ngôn ngữ phổ biến khác. Tuy nhiên, trong tiếng Việt, việc nghiên cứu độ khó văn bản còn mới mẻ và chủ yếu tập trung vào việc tìm hiểu những đặc trưng ngôn ngữ bằng các mô hình học máy thống kê.

Trong giáo dục, độ khó văn bản giúp biên soạn bài giảng, sách giáo khoa, đề thi sao cho tương đương với trình độ học sinh từng cấp độ. Để biên soạn giáo trình phù hợp với từng cấp độ, chúng ta cần xem xét đến độ khó của tất cả các từ, các câu và toàn bộ văn bản được sử dụng trong giáo trình. Vì vậy trong việc biên soạn giáo trình, chúng ta cần hết sức lưu ý đến độ khó của văn bản cũng như bài tập, câu hỏi đi kèm. Tuy nhiên, độ khó không thể xác định theo cảm quan, cảm tính của người biên soạn vì sẽ có trường hợp người biên soạn cảm thấy dễ nhưng người học ở trình độ thấp chưa chắc cảm thấy như thế. Chính vì vậy để xác định độ khó của văn bản cần có sự góp sức của khoa học máy tính. Chúng ta có thể xây dựng sẵn các kho ngữ liệu bài tập viết thuộc nhiều chủ đề khác nhau, sử dụng máy tính để phân loại tự động theo các mức độ khó khác nhau để khi cần sử dụng biên soạn giáo trình cho trình độ nào, chúng ta chỉ sử dụng những bài viết trong kho ngữ liệu đó.

Đối với các câu hỏi, phần bài tập cũng cần phải xác định độ khó. Việc lựa chọn ngữ liệu để soạn thảo câu hỏi đặc biệt là câu hỏi trắc nghiệm hoàn toàn phụ thuộc vào người soạn mà không có những tiêu chí thống nhất để đảm bảo rằng độ khó giữa các ngữ liệu nền được cung cấp trong các câu hỏi ở trình độ tương đương là thực sự tương đương. Thay vì hoàn toàn phụ thuộc vào phán đoán của từng người ra đề theo phương pháp thủ công (tức mỗi người soạn sẽ phải vừa đọc tư liệu và phán đoán sự phù hợp của từng đoạn ngữ liệu đối với mục đích kiểm

tra), rất cần có những tiêu chí khách quan, đơn giản, dễ sử dụng để hỗ trợ người soạn trong việc lựa chọn ngữ liệu. Người biên soạn cũng có thể khai thác những tài nguyên của Kho tài nguyên tiếng Việt CLC (Computational Linguistics Center: Trung tâm Ngôn ngữ học Tính toán) với từ điển đơn ngữ tiếng Việt cùng nhiều trường thông tin như hình thái, từ loại, định nghĩa, ví dụ và tần số sử dụng để xây dựng và lựa chọn những bài tập phù hợp với từng trình độ. Những kho ngữ liệu song ngữ cũng rất hữu ích trong việc xây dựng những bài tập dịch theo từng cấp độ...

Bài nghiên cứu này dưới góc độ ứng dụng đo độ khó của văn bản sẽ nhấn mạnh vào một số nội dung liên quan đến áp dụng kỹ thuật đó vào giảng dạy kỹ năng viết tiếng Việt dành cho người nước ngoài như: một số định dạng bài tập luyện viết, tần số xuất hiện một số từ vựng thông dụng và phân tích một số cấu trúc đơn giản đến phức tạp, tập trung vào các cấp độ từ A1, A2 và B1.

2. Phương pháp nghiên cứu

2.1. Cách tiếp cận

Tiếp cận theo hướng liên ngành, nhất là tính liên ngành giữa Ngôn ngữ học và Khoa học máy tính.

Tiếp cận dựa trên ngữ liệu đơn ngữ, đảm bảo một số tiêu chí về tính đại diện (ngữ liệu có nguồn uy tín, độ phổ biến cao, đa chiều: giáo trình, bài giảng đã xuất bản phục vụ giảng dạy tiếng Việt ở Việt Nam và nước ngoài), tính cân bằng (vùng miền, giới tính, thể loại...) và niên đại (khoảng 20 năm trở lại đây).

2.2. Phương pháp nghiên cứu

Phương pháp thu thập và xử lý dữ liệu: bao gồm thu thập các dữ liệu sơ cấp và thứ cấp để từng bước xây dựng và hoàn chỉnh các dạng đề thi viết với các bài tập tương ứng trình độ A1, A2, B1 theo “Khung năng lực tiếng Việt dùng cho người nước ngoài”¹. Đồng thời, vận dụng

¹ Chúng tôi đã thu thập nguồn ngữ liệu viết trình độ A1, A2, B1 từ 15 bộ giáo trình phổ biến về dạy tiếng Việt cho người nước ngoài hiện nay, có 30 dạng đề viết tương ứng với 900 bài tập. Sau khi nạp ngữ liệu, hiệu chỉnh và làm sạch ngữ liệu, huấn luyện các công cụ xử lý ngôn ngữ tự động (như WordSmith Tools, Parallel Corpus, bộ phân lớp SVM)... , kết quả nghiên cứu đã phân tích độ khó văn bản, đánh giá trình độ phù hợp với từng bài tập viết trong các giáo trình hiện có và xây dựng lại các bài tập viết này theo trình độ A1, A2, B1 (trong đó có 300 bài có trình độ A1, 300 bài trình độ A2, 300 bài trình độ B1) trên cơ sở sử dụng phần mềm thay đổi độ khó từ, câu, đoạn, bài viết.

phương pháp thống kê, phương pháp tổng hợp, phương pháp phân tích để hệ thống hóa nguồn ngữ liệu theo từng chủ đề, nội dung ở từng cấp độ cụ thể; thống kê tần suất từ xuất hiện để đánh giá độ khó và phân cấp từ dễ đến khó.

Phương pháp học máy (Machine Learning) và phương pháp học sâu (Deep Learning) nhằm tìm kiếm văn bản, phân loại văn bản, tóm tắt văn bản, dịch văn bản, phân tích độ đo, hỗ trợ dịch thuật... Phương pháp nghiên cứu dựa trên ngữ liệu (*corpus-based approach*) là cách thức thu thập, xây dựng, xử lý, khai thác các kho ngữ liệu một cách hiệu quả và đúng mục đích sử dụng. Nhóm đã xử lý ngữ liệu thu thập bằng cách làm sạch ngữ liệu, chuẩn hóa ngữ liệu, kiểm lỗi ngữ liệu, phân đoạn câu cho ngữ liệu, gán nhãn ngôn ngữ (nhãn hình thái từ, nhãn ngữ pháp từ, nhãn ngữ nghĩa từ...), từ đó huấn luyện các công cụ xử lý ngôn ngữ tự động (huấn luyện phân đoạn từ, gán nhãn tự động, chuyển đổi cây cú pháp...).

Nhóm tác giả đã sử dụng một số công cụ để xử lý ngữ liệu, như:

+ WordSmith Tools² (Concord – nhằm tìm chuỗi đồng hiện để liệt kê tất cả các sự xuất hiện của một từ hay ngữ kèm với ngữ cảnh trái và phải của từ ngữ đó; WordList – chức năng bảng từ để tạo ra một danh sách các từ có mặt trong các tập tin ngữ liệu mà chúng ta cần khảo sát như thứ hạng, tần suất, tỉ lệ phần trăm của từ trong văn bản, tỉ lệ phần trăm văn bản chứa từ đó...; và Keyword – rút trích từ khóa phục vụ tìm kiếm văn bản về chủ đề hay lĩnh vực nào đó).

+ ParaCor (Parallel Corpus)³: dùng để liệt kê, đối chiếu về hình thái, ngữ pháp, ngữ nghĩa và ngữ dụng của từ trong ngôn ngữ nguồn với ngôn ngữ đích trong một kho ngữ liệu (đơn ngữ và song ngữ). Quy trình thực hiện, gồm: nạp ngữ liệu, chỉnh sửa ngữ liệu, tìm kiếm (tìm theo từ, theo thẻ, và kết hợp tìm từ + thẻ), thống kê (tần suất xuất hiện của tất cả các từ trong ngữ liệu và hiển thị các thông tin về tần suất, tỉ lệ, độ F – tần suất được chuẩn hóa trên kích thước ngữ liệu).

+ Phương pháp so sánh độ khó của các văn bản tiếng Việt với nhau, bằng việc sử dụng bộ phân lớp SVM (Support Vector Machine: Máy học vectơ hỗ trợ). Bộ ngữ liệu sử dụng là các bài tập viết (từ giáo trình, bài giảng tiếng Việt đã xuất bản và ngữ liệu mới được xây dựng trên cơ sở các tiêu chuẩn, tiêu chí của “Khung năng lực tiếng Việt dùng cho người nước ngoài”, trình độ A1, A2, B1) đã đánh giá độ khó tương quan với nhau thông qua từ, câu, đoạn, cấu trúc, hình

² Phần mềm này do Mike Scott (Đại học Liverpool) lập trình, dựa trên phần mềm MicroConcord của Mike Scott và Tim Johns, Nxb Đại học Oxford. Xem thêm: Scott, M. (2001). *WordSmith Tools version 5.0*. Retrieved from <http://www.lexically.net/wordsmith/>

³ Công cụ này do Trung tâm Ngôn ngữ học Tính toán, Trường ĐH Khoa học Tự nhiên, ĐHQG Thành phố Hồ Chí Minh xây dựng (CLC, 2007).

thái..., làm tiền đề cho việc so sánh và lựa chọn các từ/câu/văn bản phù hợp với trình độ của người học.

3. Kết quả nghiên cứu và thảo luận

3.1 Thang đo độ khó

Để xây dựng được thang đo phù hợp, cần phải xuất phát từ đối tượng người học và yêu cầu chuẩn đánh giá năng lực tiếng Việt 6 bậc dựa trên Khung năng lực tiếng Việt dành cho người nước ngoài được ban hành theo thông tư số 17/2015/TT-BGDĐT về một số vấn đề liên quan đến các kỹ năng Nghe, Nói, Đọc và Viết. Cụ thể, kỹ năng viết được quy định chung như sau:

Bậc	Mô tả cụ thể
Bậc 1	Viết được những cụm từ, câu đơn ngắn về bản thân, gia đình, trường lớp, nơi làm việc.
Bậc 2	Viết được các mệnh đề, câu đơn nối với nhau bằng các liên từ như: <i>và, nhưng, vì</i> .
Bậc 3	Viết được đoạn, bài đơn giản, có tính liên kết về các chủ đề quen thuộc hoặc lĩnh vực mà mình quan tâm theo trật tự logic nhất định.
Bậc 4	Viết được bài chi tiết, rõ ràng về các chủ đề quen thuộc, nhiều lĩnh vực quan tâm khác nhau, biết tổng hợp, đánh giá thông tin và lập luận từ một số nguồn khác nhau.
Bậc 5	Viết được bài chi tiết, rõ ràng, bố cục chặt chẽ về các chủ đề phức tạp, biết làm nổi bật những ý quan trọng, biết mở rộng và củng cố quan điểm ở một số đoạn bằng những chứng cứ, ví dụ cụ thể và kết thúc bài viết với một kết luận phù hợp.
Bậc 6	Viết được bài rõ ràng, bố cục logic, chặt chẽ, văn phong phù hợp, trôi chảy về nhiều lĩnh vực phức tạp, giúp người đọc nhận ra những điểm quan trọng trong bài một cách dễ dàng.

Bảng 1. Mô tả về kỹ năng viết theo khung năng lực tiếng Việt dành cho người nước ngoài

Theo bảng 1, đối với các cấp độ A1, A2 và B1 tương ứng với bậc 1, bậc 2 và bậc 3, học viên được quy định về từ vựng, câu, mệnh đề, đoạn văn, bài văn khá đơn giản, có sử dụng các từ vựng có tính liên kết và được phân theo các chủ đề nhất định.

Căn cứ theo bảng mô tả cụ thể của kỹ năng viết, khi giảng dạy hay ra bài tập, đề thi, giáo viên cũng có thể lựa chọn cơ sở ngữ liệu kết hợp với phần mềm đo độ khó. Sự kết hợp đó sẽ

hướng đến sự phù hợp không chỉ trong lựa chọn bài tập, đề thi theo đúng cấp độ, mà còn hướng đến năng lực thực sự của người học.

3.2 Xử lý số liệu

Bài nghiên cứu đã kế thừa kết quả biên soạn các bài viết trong các giáo trình giảng dạy tiếng Việt cho người nước ngoài, ứng dụng phần mềm, kỹ thuật đo độ khó từ, câu, đoạn, văn bản để định dạng lại theo trình độ tương ứng với A1, A2, B1. Nhóm tác giả đã biên soạn 30 dạng đề viết tương ứng với 900 bài tập trên cơ sở sử dụng tần suất tùy từng cấp độ, tạm chia 40.000 mục từ đang có trong phần mềm, tương ứng cho 6 cấp độ trong Khung năng lực tiếng Việt dùng cho người nước ngoài để xây dựng từ, câu, đoạn, văn bản phù hợp với từng chủ đề trong Khung năng lực, “Chương trình tiếng Việt của người Việt Nam ở nước ngoài”. Sau khi hoàn thiện bước 1, chúng tôi tiếp tục đưa từ/câu/văn bản vào phần mềm đo độ khó để chạy (với phần mềm được huấn luyện – máy học), kiểm tra, đánh giá tính khách quan, phù hợp. Quá trình này phát hiện một số lỗi, buộc chúng tôi phải chỉnh sửa lại văn bản và tiếp tục chạy phần mềm cho đến khi đáp ứng yêu cầu.

Ngữ liệu	Bài viết		
	A1	A2	B1
Cấp độ	A1	A2	B1
Số lượng dạng đề	30	30	30
Số lượng bài tập tương ứng	300	300	300
Số câu tương ứng cho mỗi bài tập	05	05	05
Tổng số câu theo cấp độ	1.500	1.500	1.500
Độ dài trung bình của câu tính theo từ	10,2	15,1	19,4
Độ khó theo tần suất xuất hiện từ	7.990	10.760	15.100

Bảng 2. Thống kê ngữ liệu thu thập được sau khi xử lý

Trên cơ sở từ/câu/văn bản được biên soạn theo phần mềm, kỹ thuật đo độ khó, chúng tôi tiếp tục biên soạn các bài tập tương ứng với 30 dạng đề thi, mỗi dạng đề có 10 bài tập, mỗi bài tập có 05 câu hỏi, tổng có 4.500 câu hỏi cho 3 cấp độ, phù hợp với trình độ, năng lực theo “Khung năng lực tiếng Việt dùng cho người nước ngoài” và “Định dạng đề thi đánh giá năng lực tiếng Việt theo Khung năng lực tiếng Việt dùng cho người nước ngoài”.

3.3 Các yếu tố ảnh hưởng đến độ khó trong giảng dạy kỹ năng viết tiếng Việt

3.3.1 Yếu tố từ

Trong giảng dạy ngôn ngữ nói chung và kỹ năng viết nói riêng, từ vựng là yếu tố quan trọng và việc dạy từ vựng là việc làm tất yếu. Xác định được từ vựng phù hợp sẽ giúp phát triển tốt về cấu trúc câu, đoạn, văn bản. Tuy nhiên, đánh giá được độ khó của từ vựng để áp dụng vào quá trình giảng dạy không thể thực hiện theo cảm tính của người dạy hay người ra đề. Dựa trên một số nghiên cứu đi trước, chúng tôi xác định yếu tố ảnh hưởng tới độ khó của từ bao gồm: phương thức cấu tạo từ tiếng Việt; Nghĩa của từ và Độ thông dụng của từ [10, Tr. 17].

Yếu tố	Chỉ số về độ khó
1. Phương thức cấu tạo từ tiếng Việt	- Từ có cấu tạo đơn giản, chỉ có một âm tiết - Từ có cấu tạo phức tạp, có hai âm tiết trở lên
2. Nghĩa của từ	- Từ mang ý nghĩa từ vựng (thực từ) - Từ mang ý nghĩa ngữ pháp (hư từ)
3. Độ thông dụng của từ	- Tần suất Từ toàn dân – Từ thông dụng - Tần suất Từ Thuần Việt – Từ vay mượn - Tần suất Từ địa phương – Từ chuyên biệt - Tần suất Từ cổ – Từ mới

Bảng 3. Các chỉ số về độ khó của từ

Các chỉ số ở bảng 3 sẽ cho thấy mức độ khó của từ được diễn ra như thế nào, từ đó người biên soạn giáo trình, ra đề... có thể quan sát và đưa ra đánh giá về khối lượng kiến thức cung cấp cho người học.

Nắm được các yếu tố chỉ về độ khó của từ, chúng ta có thể kết hợp phần mềm nhằm xác định tần suất xuất hiện của từ trong phạm vi từ 10, 20, 50, 100, 200... Từ đó lựa chọn mức độ phổ biến của từ cùng việc phân tích ngữ cảnh, ngữ nghĩa và cấu trúc câu để hoàn thiện hơn cho mục đích giảng dạy.

Word	Count	Percent	F = -log(n/N)
của	14,074	2.06	1.69
tôi	13,718	2.01	1.70
một	12,107	1.77	1.75
anh	10,971	1.61	1.79
ta	10,314	1.51	1.82
không	9,442	1.38	1.86
là	8,986	1.32	1.88
đã	8,378	1.23	1.91
có	7,858	1.15	1.94
được	7,768	1.14	1.94
những	6,573	0.96	2.02
cô	6,181	0.91	2.04
cho	5,724	0.84	2.08
trong	5,367	0.79	2.10
ấy	5,147	0.75	2.12
người	4,996	0.73	2.14
ông	4,702	0.69	2.16
đó	4,573	0.67	2.17

Total word types: 16,929
Total word tokens: 682,633

Numbers
☒ All words
☐ Top 100 words

Tag
☒ No tag
☐ POS tag
☐ NER tag
☐ SEM tag

Refresh Save

Status: Completed!

Hình ảnh 1. Tần suất xuất hiện từ trong phạm vi top 100 từ đầu tiên

Từ phần mềm ngữ liệu hiện có (phần từ điển tần số tiếng Việt đầy đủ với 40.000 mục từ), chúng tôi đã truy xuất tần suất xuất hiện của 100 từ theo thứ tự từ nhỏ đến lớn, bao gồm: *của, tôi, một, anh, ta, không, là, đã, có, được, những, cô, cho, trong, ấy, người, ông, đó, ở, và, các, này, làm, sẽ, với, phải, đi, chúng tôi, vào, về, khi, để, nó, đến, sự, bị, họ, có thể, cái, ra, lại, bạn, nhiều, rất, gì, nhà, việc, con, nhưng, chúng ta, bà, mình, đang, như, nào, trên, nói, mới, cuộc, hơn, vì, trước, muốn, biết, tiên, lên, từ, sau, thì, đây, chiếc, bằng, thấy, nếu, mà, nước, điều, hai, năm, chỉ, còn, không thể, hắn, khác, ngày, lúc, rồi, nhất, rằng, nhau, tất cả, cũng, tôi, quá, cách, cả, theo, ăn, công việc, cần.*

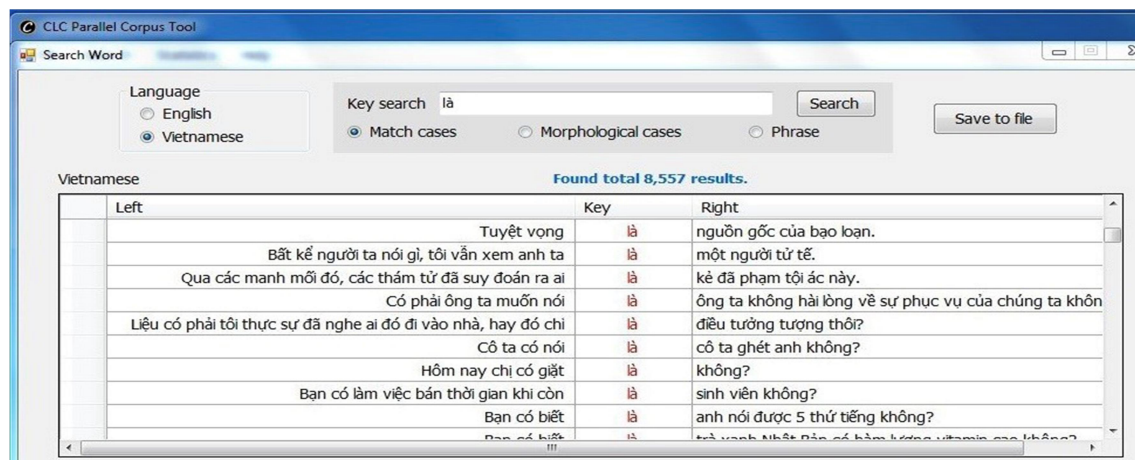
Dựa vào danh sách những từ thông dụng này, chúng ta xây dựng được vốn từ phù hợp cho người học theo trình độ để họ có thể hiểu được văn bản tiếng Việt.

3.3.2. Yếu tố câu

Theo số liệu thống kê ở trong bảng 2 ta thấy trung bình số lượng câu được phân cấp theo các cấp độ A1, A2, B1. Độ dài trung bình câu cũng sẽ tăng lên theo các cấp độ. Vì vậy, việc xác định độ khó văn bản theo những con số cụ thể là bước cơ bản để định lượng cho độ khó của một văn bản cụ thể.

Cấu trúc ngữ pháp càng đơn giản thì câu càng dễ hiểu và ngược lại. Tuy nhiên, việc xác định cú pháp đơn giản hay phức tạp phải được căn cứ trên nhiều yếu tố ngôn ngữ khác nhau như: từ loại, ngữ nghĩa, cấu trúc câu; đặc biệt là các yếu tố ngôn ngữ này có thể được xác định theo cả hướng định tính và định lượng. Chẳng hạn dưới góc độ cú pháp thì ta có thể xác định được từ loại của từ để làm rõ nghĩa của câu, nhưng điều này có thể phải kết hợp với nghĩa của từ khi từ này có thể là từ đa nghĩa hoặc đồng âm. Điều này lại có thể dẫn đến sự nhập nhằng và

khó hiểu do cấu trúc câu, do từ loại hoặc chỉ bởi do từ đồng âm. Chẳng hạn như từ “tốt” có thể là tính từ (tốt đẹp) nhưng cũng có thể là danh từ (con tốt), từ “là” có thể là hệ từ cũng có thể là động từ (là quần áo), từ “của” có thể là giới từ (của tôi) hoặc có thể là danh từ “của cải”... Vì vậy, để chọn nghĩa thông dụng cho phù hợp với trình độ người học, ta cần sử dụng danh sách đã rút ra được từ việc thống kê tần suất sử dụng của từ (theo hình ảnh 1) và vận dụng tần suất xuất hiện của từ để xác định độ khó cho câu hay đoạn văn (theo hình ảnh 2).



Hình ảnh 2. Vận dụng tần suất xuất hiện của từ để xác định độ khó câu, đoạn

3.3.3 Yếu tố văn bản

Độ khó văn bản không chỉ được đánh giá dựa vào sự ảnh hưởng của các yếu tố từ và yếu tố câu đối với văn bản, mà còn được đánh giá dựa vào sự ảnh hưởng của các yếu tố ngôn ngữ trong một chỉnh thể của văn bản. Số lượng câu trong một văn bản luôn được xem xét như là một trong những yếu tố quan trọng để đánh giá độ khó văn bản. Số lượng câu đơn và số lượng câu ghép được sử dụng trong một văn bản theo cấp độ cho thấy đa số chúng đều có chiều hướng tỉ lệ thuận đối với độ khó văn bản theo cấp độ, nghĩa là ở cấp độ thấp có khuynh hướng dùng câu đơn, và tỉ lệ câu ghép tăng dần khi cấp độ cao hơn. Văn bản có nhiều câu ghép, độ sâu của cây cú pháp của các câu trong văn bản tăng, thì độ phức tạp của kết cấu văn bản tăng và độ khó của văn bản cũng tăng theo [3, Tr. 551].

Việc xác định yếu tố ngôn ngữ ảnh hưởng đến độ khó văn bản về mặt kết cấu văn bản được xem xét không chỉ về yếu tố ngôn ngữ bề mặt như số lượng câu, số lượng câu đơn, số

lượng câu ghép... mà còn về các mối liên kết trong câu và toàn bộ văn bản. Chính vì vậy, các yếu tố ngôn ngữ chỉ ra các mối liên kết trong văn bản luôn được chú trọng trong các nghiên cứu về độ khó văn bản.

Trên cơ sở thống kê, lựa chọn nguồn ngữ liệu từ các công trình đã công bố, bài nghiên cứu đã trích xuất các dạng bài tập Viết tương ứng với trình độ A1, A2, B1 để đưa vào phần mềm xử lý, phân loại độ khó của từ, câu, đoạn, văn bản thông qua từ vựng (tần suất hình ảnh 1), độ phức tạp của cấu trúc câu (số lượng mệnh đề, chiều dài câu), tổ chức văn bản (liên kết văn bản), hình thức văn bản... Kết quả phân tích cho thấy, những giáo trình, bài giảng, đề thi đánh giá năng lực tiếng Việt dùng cho người nước ngoài hiện nay chưa đảm bảo phân cấp độ phù hợp theo các tiêu chuẩn, tiêu chí của “Khung năng lực tiếng Việt dùng cho người nước ngoài” và “Định dạng đề thi đánh giá năng lực tiếng Việt theo Khung năng lực tiếng Việt dùng cho người nước ngoài”. Nhiều bài đọc mang tính chủ quan, cảm tính, sử dụng các từ, liên từ, đoạn văn bản có độ khó không phù hợp với trình độ (gồm cả hai chiều, độ khó thấp nhưng trình độ cao và độ khó cao nhưng thực tế ngữ liệu có trình độ thấp).

3.4 Một số dạng bài tập có thể sử dụng trong các đề thi viết trình độ A1, A2, B1

Từ những kết quả nghiên cứu về ứng dụng phần mềm để đo độ khó được làm rõ phía trên và theo kinh nghiệm giảng dạy tiếng Việt cho người nước ngoài của tác giả trong vòng 12 năm nay cùng với quá trình khảo sát từ các giáo trình đã sử dụng. Chúng tôi xin được liệt kê một số dạng bài tập/đề thi về kỹ năng viết thường được sử dụng, những dạng bài tập này sẽ là tiền đề phục vụ cho quá trình giảng dạy tiếng Việt, biên soạn giáo trình cũng như ra đề thi năng lực tiếng Việt cho người nước ngoài, đặc biệt là về kỹ năng viết. Cụ thể như sau:

Stt	Dạng bài tập/đề	Stt	Dạng bài tập/đề
1	Sắp xếp từ thành câu hoàn chỉnh/đúng	16	Dùng gợi ý để trả lời câu hỏi
2	Sắp xếp câu thành hội thoại đúng	17	Thêm từ cùng loại vào dãy từ còn thiếu
3	Hoàn thành hội thoại	18	Chọn từ vựng không cùng nhóm
4	Điền thanh điệu vào từ, câu	19	Chọn từ thích hợp điền vào chỗ trống
5	Đọc và viết chính tả	20	Chọn câu thích hợp điền vào chỗ trống
6	Hoàn thành câu theo gợi ý cho sẵn	21	Tìm từ gần nghĩa
7	Dùng kết cấu để hoàn thành câu	22	Tìm từ trái nghĩa

8	Nối từ ở cột A tương ứng với cột B	23	Dùng các từ vựng cho sẵn viết thành các đoạn văn ngắn
9	Nối câu hỏi ở cột A tương ứng với câu trả lời ở cột B	24	Chuyển câu khẳng định thành câu phủ định hoặc nghi vấn
10	Viết lại câu sao cho nghĩa không đổi	25	Tìm lỗi sai trong đoạn văn và sửa lại cho đúng
11	Chọn từ đúng sau đó viết lại câu	26	Tìm lỗi sai trong câu và sửa lại cho đúng
12	Đặt câu hỏi cho các câu trả lời	27	Dùng cấu trúc để liên kết câu
13	Đặt câu hỏi cho từ/cụm từ được gạch chân	28	Viết thư
14	Đặt câu với từ cho sẵn	29	Trò chơi ô chữ
15	Ghép từ vựng với tranh/Xem tranh và điền từ	30	Viết luận theo chủ đề

Bảng 4. Một số dạng đề/bài tập kỹ năng viết tiếng Việt cho người nước ngoài

Trong quá trình biên soạn thì câu hỏi và phần từ vựng cần được xác định độ khó. Nếu có từ khó hay cấu trúc khó, phần mềm sẽ cảnh báo để người biên soạn có thể thay đổi bằng cách lựa chọn một từ đồng nghĩa hay một cấu trúc ngữ pháp khác sao cho có độ khó thấp hơn. Đây là một trong những giá trị cốt lõi mà kỹ thuật đo độ khó mang lại. Căn cứ vào 30 dạng bài tập này, nhóm tác giả đã tổng hợp được một khối ngữ liệu viết vô cùng đồ sộ với khoảng 4.500 câu hỏi nhằm phục vụ cho việc giảng dạy tiếng Việt cho người nước ngoài.

4. Kết luận và kiến nghị

Nghiên cứu về độ khó văn bản có ý nghĩa trong việc biên soạn các tài liệu giảng dạy ngoại ngữ hay xây dựng ngân hàng đề thi Đánh giá năng lực sử dụng ngoại ngữ. Việc xác định các yếu tố ngôn ngữ từ cấp độ từ, câu và văn bản có sự ảnh hưởng đến độ khó văn bản với

những con số cụ thể có thể phục vụ trong quá trình giảng dạy ngoại ngữ như chọn lọc từ vựng, văn bản, viết văn bản mẫu, biên soạn khối ngữ liệu thô, từ điển...

Kết quả nghiên cứu đã chỉ ra việc sử dụng ngữ liệu áp dụng vào 30 dạng bài tập có thể sử dụng trong giảng dạy hoặc đề thi kỹ năng viết tương ứng với 900 bài tập với khoảng 4.500 câu hỏi được soạn mới theo tiêu chuẩn, tiêu chí, chủ đề, yêu cầu của “Khung năng lực tiếng Việt dùng cho người nước ngoài”, “Chương trình tiếng Việt cho người Việt Nam ở nước ngoài”, “Định dạng đề thi đánh giá năng lực theo Khung năng lực tiếng Việt dùng cho người nước ngoài” và ứng dụng phần mềm, kỹ thuật đo độ khó đảm bảo tính phù hợp, định lượng, khách quan, khoa học.

Kiến nghị: Cần xây dựng, phát triển khối ngữ liệu viết lớn hơn và ứng dụng thêm nhiều kỹ thuật, phần mềm của khoa học máy tính để đo độ khó, phân loại cấp độ theo “Khung năng lực tiếng Việt dùng cho người nước ngoài” nhằm phục vụ giảng dạy, biên soạn giáo trình, bài giảng và quỹ đề thi đánh giá năng lực phù hợp, thống nhất.

TÀI LIỆU THAM KHẢO

1. An-Vinh Luong, Diep Nguyen, Dien Dinh (2017), “Examining the Text-length Factor in Evaluating the Readability of Literary Texts in Vietnamese Textbooks”; *9th International Conference on Knowledge and Systems Engineering (KSE), Hue, Vietnam, 2017*, pp. 36-41. DOI: 10.1109/KSE.2017.8119431.
2. Đinh Điền (2018), *Ngôn ngữ học ngữ liệu*, Nxb ĐHQG Tp. Hồ Chí Minh.
3. Nguyễn Thị Như Diệp, Trần Thị Phương Lý (2022), “Ứng dụng độ khó của văn bản trong việc giảng dạy ngôn ngữ”, *Tạp chí khoa học trường Đại học Quốc tế Hồng Bàng*, số đặc biệt 12/2022.
4. Nguyễn Văn Huệ (cb) (2001, 2004, 2013), *Giáo trình tiếng Việt dành cho người nước ngoài (quyển 1, 2, 3, 4)*; Nxb Giáo dục và Nxb ĐHQG Tp. Hồ Chí Minh.
5. Nguyễn Việt Hương (2013), *Tiếng Việt cơ sở dành cho người nước ngoài (quyển 1, 2)*, Nxb ĐHQG Hà Nội.
6. Nguyễn Việt Hương (2013), *Tiếng Việt nâng cao dành cho người nước ngoài (quyển 1, 2)*, Nxb ĐHQG Hà Nội.

7. Khoa Việt Nam học và Tiếng Việt, Trường Đại học Khoa học Xã hội và Nhân văn, ĐHQG Hà Nội (2019), *Tiếng Việt cho người nước ngoài (tập 1, 2, 3, 4)*, Hà Nội.
8. Tony Mc, A. W. (2001), *Corpus Linguistics*; Sil. (2017), *Langguage Repot*, Retrieved from <https://www.ethnologue.com>
9. Viện Việt Nam học và Khoa học phát triển, ĐHQG Hà Nội (2014), *Tiếng Việt trình độ A, B, C*, Nxb ĐHQG Hà Nội.
10. Nguyễn Thị Thanh Vân (2022), “Độ khó của từ vựng trong các giáo trình dạy tiếng Việt cho người nước ngoài”, *Tạp chí ngôn ngữ và đời sống*, số 4(324)-2022.