

Truyện ngắn Lý Lan và Y Ban từ góc nhìn lí thuyết độ khó của văn bản

Nguyễn Thị Như Diệp^{1,*}, Trịnh Khánh Linh^{2,3,**}

¹Học Viện Hàng không Việt Nam, ²Trường Đại học Nguyễn Tất Thành, ³Trường Đại học Sài Gòn
*diepntn@vaa.edu.vn; **linhtk@ntt.edu.vn

Tóm tắt

Từ lâu, trên thế giới, độ khó của văn bản đã được nghiên cứu trong cả lĩnh vực ngôn ngữ học ứng dụng và ngôn ngữ học tính toán; trong đó, các nghiên cứu chủ yếu cho tiếng Anh và một số ngôn ngữ khác với nhiều kết quả ứng dụng cao. Bài viết phần lớn sử dụng phương pháp nghiên cứu định lượng nhằm khảo sát các yếu tố ngôn ngữ ảnh hưởng đến độ khó của truyện ngắn Lý Lan và Y Ban từ góc nhìn lí thuyết độ khó của văn bản. Qua khảo sát và phân tích bộ ngữ liệu gồm 62 truyện ngắn của Lý Lan và Y Ban, cho thấy, cấp độ khó truyện ngắn Lý Lan và Y Ban ở mức cấp độ khó “trung bình”. Từ kết quả nghiên cứu, bài viết đưa ra các đề xuất nhằm sử dụng ngữ liệu ở các lớp, cấp học phù hợp; cập nhật vào kho ngữ liệu tiếng Việt nghiên cứu về độ khó của văn bản; đề xuất kết hợp khung lí thuyết này vào ứng dụng thực tiễn giảng dạy, biên soạn giáo trình và xác định đặc trưng phong cách tác giả ở góc nhìn định lượng.

Nhận 05/09/2023
Được duyệt 26/10/2023
Công bố 01/11/2023

Từ khóa

Độ khó của văn bản,
truyện ngắn,
yếu tố ngôn ngữ,
cấp độ khó,
ứng dụng thực tiễn

© 2023 Journal of Science and Technology - NTTU

1 Đặt vấn đề

Trên thế giới, độ khó của văn bản (ĐKVB) rất được quan tâm và nghiên cứu từ cuối thế kỷ 19 vì tính ứng dụng cao. Tuy nhiên, các nghiên cứu về vấn đề này đa số chỉ tập trung vào văn bản là tiếng Anh và một số ngôn ngữ khác [1,2,3]. Tại Việt Nam, hướng nghiên cứu về ĐKVB Tiếng Việt còn bỏ ngõ và vẫn là “ngôn ngữ nghèo tài nguyên” [4] nhất là với ngữ liệu truyện ngắn. Những năm gần đây, các nghiên cứu đã bước đầu đề xuất giải pháp giúp cho người giảng dạy, biên tập lựa chọn tài liệu và xây dựng ngân hàng đề [3] hay công trình chuyên biệt đưa ra lý thuyết đầu tiên về việc đo ĐKVB tiếng Việt với các tiêu chí để đánh giá độ khó văn bản, mô hình hóa các công cụ đo độ khó tự động [5]. Việc nghiên cứu ĐKVB có đóng góp quan trọng và tiềm năng ứng dụng không chỉ trong các phân ngành của ngôn ngữ học mà còn các liên ngành khác có liên quan đến ngôn ngữ: ngôn ngữ học ngữ liệu, ngôn ngữ học ứng dụng, sử dụng trí tuệ nhân tạo (AI) trong xử lý ngôn ngữ... và ứng dụng các kết quả nghiên cứu theo

hướng ĐKVB trong thực tiễn giảng dạy ngôn ngữ, nghiên cứu lý luận và phê bình văn học.

Trong khuôn khổ và giới hạn, bài viết sử dụng ngữ liệu là các văn bản trích từ các tập truyện ngắn của hai tác giả Lý Lan và Y Ban. Lý Lan và Y Ban là hai trong số các nhà văn nữ đã có những đóng góp quan trọng cho sự phát triển, đổi mới về nội dung và tư duy trong truyện ngắn Việt Nam ở giai đoạn văn học đương đại. Các sáng tác của hai nhà văn được đón nhận rộng rãi trên văn đàn, trong lòng độc giả và được nhiều nhà xuất bản tìm đến. Bên cạnh đó, sáng tác của Lý Lan cũng được đưa vào chương trình giảng dạy Ngữ văn 7 với bản văn “*Cổng trường mở ra*”.

Tuy nhiên, hiện nay có rất ít các công trình nghiên cứu chuyên biệt về truyện ngắn của Lý Lan và Y Ban. Đối với nhà văn Lý Lan, chủ yếu là các bài giới thiệu sách, phỏng vấn trên báo. Hiện tại chỉ có một nghiên cứu tập trung xác định, lý giải đặc trưng nội dung và nghệ thuật truyện ngắn của nhà văn Lý Lan [6]. Đối với tác giả Y Ban, một số nghiên cứu tìm hiểu về đặc điểm nổi bật

trong sáng tác [7], cách thể hiện vấn đề nữ quyền trong nội dung [8] cũng như những đóng góp vào văn xuôi đương đại Việt Nam [9]. Đặc biệt, chưa có công trình nghiên cứu truyện ngắn của hai tác giả này dưới góc nhìn lý thuyết ĐKVB.

Chính vì vậy, việc khảo sát các yếu tố ngôn ngữ (YTNN) nhằm xác định cấp độ khó của truyện ngắn Lý Lan và Y Ban từ góc nhìn lý thuyết ĐKVB là cần thiết. Từ kết quả nghiên cứu, có thể đề xuất các truyện ngắn thích hợp dùng cho các lớp, cấp học tương ứng; bổ sung vào kho ngữ liệu tiếng Việt phục vụ nghiên cứu ĐKVB; đề xuất ứng dụng vào thực tiễn trong việc chọn lọc văn bản phục vụ cho giảng dạy tiếng Việt, biên soạn giáo trình, sách giáo khoa và xác định đặc trưng phong cách tác giả ở góc nhìn định lượng của lý thuyết ĐKVB.

Có nhiều quan điểm về khái niệm văn bản, có quan điểm cho rằng “văn bản là bản ghi bằng chữ viết hoặc chữ in một phát ngôn hoặc một thông báo bằng ngôn ngữ; đây là đơn vị nhỏ nhất có tính thống nhất và tính độc lập của giao tiếp bằng ngôn ngữ” [10]. Lại có quan điểm cho rằng, văn bản là “một phần của ngôn ngữ có những mục đích có thể nhận diện được. Việc này dẫn đến việc phân loại chúng thành một số loại khác nhau về mục đích và đặc trưng ngôn ngữ” [11]. Trong một nghiên cứu về ĐKVB đối với tiếng Việt (có đối chiếu với tiếng Anh) đưa ra định nghĩa “văn bản là sản phẩm của tạo lời, tồn tại dưới dạng thức viết, có tính thống nhất về nội dung và hình thức nhằm thực hiện chức năng thông báo thông tin” [5].

Từ cuối thế kỷ 19, thế giới đã ghi nhận các thành tựu nghiên cứu về ĐKVB được ứng dụng vào thực tiễn trong nhiều lĩnh vực. Công trình đầu tiên về ĐKVB dựa trên 3 yếu tố để đánh giá độ khó của từ vựng bằng công thức đếm trên 1.000 từ và tính chỉ số trung bình của các từ, đưa ra phương pháp dùng để đo lường từ vựng khó trong hơn 11 sách giáo khoa có mức độ từ vựng khó khác nhau [12]. Đến năm 1935, nghiên cứu được xem là kinh điển về YTNN ảnh hưởng đến ĐKVB với mục tiêu tìm ra các tiêu chí mang tính khách quan để phát triển công thức có thể đo lường ĐKVB [13], hay nghiên cứu vào năm 1969 xem xét mối quan hệ giữa độ dài câu liên quan đến số lượng từ có sự tương ứng với nhau qua phương trình hồi quy [14]. Ngoài ra, còn có một số công cụ đo ĐKVB như công thức đo ĐKVB Flesch [15] tích hợp vào chương trình Microsoft Word nhằm mục đích xác định ĐKVB với thang điểm đánh giá khả năng đọc hiểu; trong đó, ý nghĩa của thang điểm đánh

giá là giá trị điểm càng cao thì văn bản càng dễ hiểu và ngược lại; hay công cụ Coh-Metrix [16] được tích hợp trên trang mạng, người dùng có thể truy cập vào sử dụng. Bên cạnh đó, việc tích hợp các công thức đo ĐKVB Flesch, Coh-Metrix indices và Lexile measure trong khung tham chiếu ngoại ngữ chung của Châu Âu nhằm phân cấp trình độ của người học theo hệ thống từ lúc bắt đầu cho đến mức tương đương như người bản ngữ. Đối với các nghiên cứu về ĐKVB tiếng Việt tại Việt Nam, từ 1983 có nghiên cứu được công bố của cùng một nhóm tác giả khi đề xuất công thức đo ĐKVB đầu tiên dành cho tiếng Việt [17] đến hiện tại có một số nghiên cứu đã sử dụng thêm các phần mềm hỗ trợ để xử lý văn bản tiếng Việt do Trung tâm Ngôn ngữ học Tính toán phát triển dựa như hệ thống máy học (YamCha) với các nhóm công cụ như: công cụ tách từ tiếng Việt (CLC_VN_WS), công cụ gán nhãn từ loại tiếng Việt (CLC_VN_POS), công cụ gán nhãn NE tiếng Việt (CLC_VN_NER) và công cụ chuẩn hóa tập tin (CLC_VN_Toolkit). Qua khảo lược, có thể kể đến các nghiên cứu có ứng dụng các công cụ đo ĐKVB như: nghiên cứu áp dụng ĐKVB trong việc xây dựng ngữ liệu giáo trình tiếng Việt cho người nước ngoài [18], khảo sát các YTNN ảnh hưởng đến ĐKVB - ngữ liệu TOEIC, cấp độ cơ bản [3], so sánh các văn bản tiếng Việt theo độ khó [19] hay trong nghiên cứu về các YTNN ảnh hưởng đến ĐKVB tiếng Việt (có đối chiếu với tiếng Anh) [5] và việc xây dựng mô hình đánh giá ĐKVB tiếng Việt [20].

Tương tự như khái niệm về văn bản, từ các góc nhìn khác nhau thì ĐKVB cũng được định nghĩa khác nhau. ĐKVB được hiểu theo nghĩa chung nhất là “tổng hợp các yếu tố (bao gồm cả các yếu tố tương tác) ảnh hưởng đến khả năng đọc - hiểu văn bản của một nhóm đối tượng cụ thể với tốc độ đọc bình thường và thoải mái” [21]. Bên cạnh đó, còn có định nghĩa ĐKVB “là độ đo xác định xem một văn bản là dễ hay khó đọc như thế nào. ĐKVB phụ thuộc nhiều vào các YTNN như cách dùng từ, ngữ, câu, phong cách của văn bản...” [22]. Như vậy, văn bản (VB) có thực hiện được chức năng thông tin của nó hay có sự phù hợp với người tiếp nhận sẽ phụ thuộc vào hai nhân tố chính đó là “Độ hiểu về văn bản” (Comprehensibility) là khả năng hiểu văn bản của người đọc và “Độ phức tạp của văn bản” (Complexibility) là liên quan đến các yếu tố nội tại của chính VB đó. “Độ phức tạp của văn bản” (ĐPTVB) bao gồm hai phương diện đó là: “Độ rõ của văn bản” (Text Legibility/ Legibility) tập trung các yếu tố về

hình thức trình bày VB và “Độ khó của văn bản” (Readability/ Text Readability/ Text Difficulty/ Textual Difficulty) tập trung hướng nghiên cứu về các yếu tố nội dung của VB, các YTNN bề mặt và YTNN bề sâu thuộc nội dung VB viết [5].

Như vậy có thể thấy, việc ứng dụng lý thuyết vào thực tiễn của các nghiên cứu về ĐKVB từ lâu đã trở nên rất quan trọng và phổ biến [3,5,15,16,19,20] vì những lợi ích mà nó mang lại đối với nhu cầu thực tế của xã hội trong các lĩnh vực từ giáo dục, văn hóa, xã hội đến chính trị, kinh tế và các lĩnh vực khác.

Bài viết đi sâu vào giải quyết các câu hỏi nghiên cứu như sau: (1) ở cấp độ Từ, YTNN nào là yếu tố ảnh hưởng cao đến ĐKVB trong truyện ngắn của Lý Lan và Y Ban; (2) ở cấp độ Câu, YTNN nào là yếu tố ảnh hưởng cao đến ĐKVB trong truyện ngắn của Lý Lan và Y Ban; và (3) truyện ngắn của Y Ban và Lý Lan được xác định ở cấp độ khó nào? Trên cơ sở đó, bài viết sẽ đưa ra một số đề xuất ứng dụng thực tiễn.

2 Đối tượng, dữ liệu và phương pháp nghiên cứu

- Đối tượng nghiên cứu: các YTNN phục vụ giải đáp cho câu hỏi nghiên cứu của bài viết cụ thể là: (1) từ Hán Việt (tỉ lệ), (2) trung bình độ dài câu tính bằng từ, (3) tổng số lượng từ trong một VB, và (4) tổng số lượng câu trong một VB.

- Dữ liệu nghiên cứu: 62 truyện ngắn đã được xuất bản của Lý Lan và Y Ban cụ thể là 22 truyện trong tuyển tập *Hồi xuân* [23], 09 truyện trong tuyển tập *Truyện ngắn 4 cây bút nữ* [24] và 31 truyện trong tuyển tập *Truyện ngắn Y Ban* [25].

- Dữ liệu sơ cấp được thu thập từ quá trình xử lý thủ công kết hợp sử dụng công cụ CLC_VN_Toolkit [26,27] được huấn luyện để chuẩn hóa các lỗi chính tả của tập tin (dấu câu, i-y, dấu thanh, unicode dựng sẵn), tách câu, tách từ, gán nhãn từ loại.

- Phương pháp nghiên cứu và phân tích dữ liệu: Sử dụng phương pháp nghiên cứu định lượng qua việc sử dụng công cụ CLC_VN_Toolkit kết hợp thống kê mô tả, tổng hợp kết quả khảo sát, từ đó xác định những yếu tố Từ, Câu và Văn bản ảnh hưởng đến độ khó của truyện ngắn của Lý Lan và Y Ban nhằm giải quyết câu hỏi nghiên cứu đã đặt ra. Số liệu được tính toán xác định ĐKVB các truyện ngắn (mức độ và cấp lớp) đối với thang đo ĐKVB tiếng Việt theo công thức [17]:

$$(\text{Độ khó của văn bản}) RL = 0,27 \times WD + 0,13 \times SL + (\text{Độ dài câu tính bằng từ}) + 1,74$$

Trong đó: WD, SL được xác định như sau:

WD (Word Difficulty: chỉ số độ khó từ vựng) = *Số từ Hán Việt/Tổng số lượng từ trong VB*

SL (độ dài trung bình của câu) = *Tổng số từ/Tổng số câu*

Như vậy, kết quả thu được sau khi áp dụng công thức ĐKVB tiếng Việt sẽ được xác định theo tiêu chí đánh giá mức độ khó của thang đo như sau:

Bảng 1 Bảng quy ước mức độ khó văn bản tiếng Việt

Loại	Mức độ	Cấp lớp	Độ khó (RL)
1	Rất dễ	1-2	1,0 -2,0
2	Dễ	3-4	2,2-4,0
3	Dễ vừa phải	5-6	4,2-6,0
4	Tiêu chuẩn	7-8	6,2-8,0
5	Khó vừa phải	9-10	8,2-10,0
6	Khó	11-12	10,2-12,0
7	Rất khó	Đại học	12,2 trở lên

3 Kết quả và thảo luận

Khi đánh giá ĐKVB, các nghiên cứu xem xét rất nhiều YTNN cả về mặt định lượng lẫn định tính để phân tích ở 3 cấp độ chính của ngôn ngữ là: Từ - Câu - Văn bản, chẳng hạn như các yếu tố: độ dài từ, từ loại, nguồn gốc từ, tần suất từ, độ dài câu, cấu trúc câu, độ sâu cây cú pháp, các yếu tố liên kết văn bản, thể loại văn bản, v.v... [1, 5, 27-29]. Bài viết chỉ trình bày, phân tích định lượng các YTNN đã được đề cập trong câu hỏi nghiên cứu: (1) yếu tố “từ Hán Việt” trong cấp độ “Từ”, (2) yếu tố “trung bình độ dài câu tính bằng từ” trong cấp độ “Câu”, (3) yếu tố “tổng số lượng từ trong một văn bản”, và “tổng số lượng câu trong một văn bản” ở cấp độ “Văn bản”.

3.1 Yếu tố từ Hán Việt

Từ Hán Việt có mức độ ảnh hưởng cao hơn đến ĐKVB khi so sánh với các YTNN khác trong cấp độ “Từ”. Trong quá trình phát triển của tiếng Việt, từ Hán Việt trở thành một phần không thể thiếu trong cấu trúc ngữ pháp của tiếng Việt khi hơn 60% vốn từ tiếng Việt là gốc Hán. Số lượng từ Hán Việt chiếm tỉ lệ lớn trong từ vựng tiếng Việt. Các từ Hán Việt trong lĩnh vực khoa học kĩ thuật được đánh giá là khó hiểu hơn từ thuần Việt [27-32]. Việc sử dụng từ Hán Việt giúp VB trang trọng hơn nhưng cũng làm tăng độ khó của văn bản bởi đặc trưng lớp ý nghĩa trừu tượng, đòi hỏi người đọc phải có kiến thức và cả ngữ cảnh để hiểu rõ ý nghĩa của từ. Trong nghiên cứu xác định vai trò của từ Hán Việt

khí đánh giá ĐKVB nếu kết hợp yếu tố này với các YTNN bề mặt (trung bình độ dài từ tính bằng kí tự, tính bằng chữ...) thì độ chính xác sẽ tăng cao và ngược lại không có sẽ có độ chính xác thấp [3].

Nghiên cứu tiến hành thống kê từ Hán Việt được sử dụng trong truyện ngắn của Lý Lan và Y Ban, trên cơ sở đó thực hiện việc tính chỉ số độ khó từ vựng (Word Difficulty: WD) còn gọi là tỉ lệ từ Hán Việt [27]. Kết quả này được trình bày cụ thể trong bảng sau:

Bảng 2 Từ Hán Việt trong truyện ngắn của Lý Lan

TT	Tên Truyện ngắn	Tổng Số lượng từ	Từ Hán – Việt		Số lượng từ phân biệt	Từ Hán – Việt phân biệt	
			Số từ Hán Việt	Tỉ lệ (%)		Số từ Hán - Việt phân biệt	Tỉ lệ (%)
1	Biển trong mưa	2.222	471	21,20	624	161	25,80
2	Tháng chạp	2.405	485	20,17	853	246	28,84
3	Cuối tuần	1.722	355	20,62	667	170	25,49
4	Lắp ghép hạnh phúc	1.842	382	20,74	597	189	31,66
5	Con mèo tưởng đã đi xa	1.600	290	18,13	539	140	25,97
6	Bay qua bầu trời thành phố đêm giao thừa	1.564	356	22,76	614	168	27,36
7	Đêm thảo nguyên	2.592	556	21,45	835	281	33,65
8	Vườn hoàng tử nhỏ	1.410	292	20,71	497	120	24,14
9	Mẹ và con	2.596	598	23,04	791	223	28,19
10	Mơ ước hoang đường	1.363	248	18,20	550	136	24,73
11	Pha lê	1.738	387	22,27	701	200	28,53
12	Bông vạn thọ	1.499	296	19,75	606	161	26,57
13	Tâm hồn	1.483	424	28,59	531	156	29,38
14	Hồi xuân	1.860	419	22,53	824	237	28,76
15	Romeo và Juliet	1.377	289	20,99	626	172	27,48
16	Hạnh phúc chon kinh	1.727	469	27,16	697	189	27,12
17	Phương pháp hiện thực	1.399	280	20,01	598	153	25,59
18	Tình chỉ đẹp	1.454	331	22,76	673	185	27,49
19	Cẩm Sơn	1.810	342	18,90	680	184	27,06
20	Cảm giác	1.060	245	23,11	462	108	23,38
21	Lòng Hồ	1.602	339	21,16	651	168	25,81
22	Phi trường Đài Bắc	3.562	734	20,61	1.050	258	24,57
23	Xuân thì	2.070	485	23,43	753	195	25,90
24	Kết thúc có hậu	939	222	23,64	453	128	28,26
25	Núi không	2.180	497	22,80	798	242	30,33
26	Tại sao anh làm điều đó	1.493	316	21,17	605	173	28,60
27	Bé xuân về	1.042	215	20,63	502	117	23,31
28	Đường dài hạnh phúc	2.502	522	20,86	888	258	29,05
29	Truyện ma	1.343	301	22,41	578	150	25,95
30	Một đời	1.008	195	19,35	479	122	25,47
31	Bạn	904	176	19,47	400	83	20,75
Tổng		53.368	11.517	21,58	5.473	20.122	27,20

Số liệu thống kê trong Bảng 2 cho thấy, từ Hán Việt sử dụng trong truyện Lý Lan chiếm tỉ lệ trung bình khoảng 21,58%. Tỉ lệ từ này dao động từ tối thiểu 18,13% đến tối đa 28,59%. Như vậy tỉ lệ từ Hán Việt trong 31 truyện không vượt quá 30%. Để tránh sự lặp lại của một từ Hán Việt trong một văn bản, bài viết xem xét yếu tố “từ Hán Việt phân biệt”, kết quả cho thấy tỉ lệ từ này

được sử dụng dao động từ tối thiểu 20,75% đến tối đa là 33,65%. Thêm vào đó, số liệu còn cho thấy trong 31 truyện thì chỉ có 3 truyện sử dụng số lượng từ Hán Việt phân biệt có tỉ lệ trên 30% là: *Núi không, Lấp ghép hạnh phúc* và *Đêm thảo nguyên*. Và tỉ lệ trung bình số lượng từ Hán Việt phân biệt sử dụng trong 31 truyện Lý Lan ở mức 27,20%.

Bảng 3 Từ Hán Việt trong truyện ngắn của Y Ban

TT	Tên Truyện ngắn	Tổng số lượng từ	Từ Hán Việt		Tổng số lượng từ phân biệt	Từ Hán – Việt phân biệt	
			Số từ Hán Việt	Tỉ lệ (%)		Số từ Hán - Việt phân biệt	Tỉ lệ (%)
1	Bức thư gửi mẹ Âu Cơ	6.977	1.418	20,32	1.488	434	29,17
2	Bồ công anh nở bên hồ nước trong	8.493	1.708	20,11	1.492	474	31,77
3	Có thể có và có thể không	7.823	1.689	21,59	1.721	546	31,73
4	Con yêu tinh	6.629	1.467	22,13	1.697	479	28,23
5	Gà ấp bóng	5.003	974	19,47	1.114	330	29,62
6	I am đàn bà	7.237	1.772	24,49	1.451	387	26,67
7	Người đàn bà đứng trước gương	3.009	547	18,18	872	256	29,36
8	Sự nhầm lẫn bò cái	3.339	713	21,35	974	289	29,67
9	Trong khu vườn nghệ sỹ	3.311	609	18,39	755	166	21,99
10	Cắm cù	13.757	2.952	21,46	2.295	702	30,59
11	Cái Tý	4.941	967	19,57	1.112	266	23,92
12	Anh Quảng	4.324	845	19,54	1.261	329	26,09
13	Gái góa là gái góa ơi	11.615	2.280	19,63	2.135	539	25,25
14	Chín khúc hát ru của thiên thần	4.092	865	21,14	783	220	28,10
15	Đôi găng tay da màu nâu	2.655	472	17,78	733	202	27,56
16	Xích lô	2.739	488	17,82	780	189	24,23
17	Con gái mang cuộc đời mẹ	5.714	1.033	18,08	1.201	321	26,73
18	Jô	2.938	527	17,94	756	223	29,50
19	Làng Cò	6.812	1.435	21,07	1.703	484	28,42
20	Chồng tôi	2.029	425	20,95	712	208	29,21
21	Người đàn bà có ma lực	8.940	1.710	19,13	1.807	531	29,39
22	Quê nội	3.072	664	21,61	925	245	26,49
23	Bản lý lịch tự thuật	3.285	780	23,74	952	293	30,78
24	Vùng sáng ký ức	5.599	1.143	20,41	1.269	340	26,79
25	Không đề	3.150	648	20,57	955	314	32,88
26	Chợ rằm dưới gốc dâu cổ thụ	4.475	755	16,87	975	228	23,38
27	Con quỷ nhỏ trong tôi	6.879	1.539	22,37	1.300	438	33,69
28	Em vẫn gọi tên anh là nước Nga	3.349	686	20,48	991	302	30,47
29	Mẹ không thể xin lỗi con	4.370	842	19,27	993	278	28,00
30	Mỗi người đàn ông chỉ của riêng một người đàn bà	5.034	906	18,00	1.136	323	28,43
31	Ồi những chú ngựa bắt kham	1.851	419	22,64	631	192	30,43
Tổng		163.441	33.278	20,36	36.969	10.528	28,48

Số liệu thống kê trong bảng 3 cho thấy số lượng từ Hán Việt xuất hiện trong ngữ liệu dao động từ 16,87% đến tối đa là 24,49%. Kết quả sử dụng số lượng từ Hán Việt trong 31 truyện ngắn của Y Ban có tỉ lệ không vượt quá 25% và trung bình tỉ lệ số lượng từ Hán Việt được sử dụng là 20,36%. Bên cạnh đó, tỉ lệ trung bình số lượng từ Hán Việt phân biệt thống kê được chiếm 28,48%. Trong đó, có 8 truyện ngắn trong 31 truyện có tỉ lệ lượng từ Hán Việt phân biệt sử dụng trên 30% là: *Ơi những chú ngựa bất kham, Em vẫn gọi tên anh là nước Nga, Cầm cù, Bản lý lịch tự thuật, Có thể có và có thể không, Bò công anh nở bên hồ nước trong, Không đề, Con quý nhỏ trong tôi*. Số liệu cho thấy tỉ lệ số từ Hán Việt phân biệt dao động tối thiểu từ 21,99% đến 33,69%.

Bên cạnh đó, trong công trình nghiên cứu ĐKVB tiếng Việt với khối lượng từ mang tính tổng quát hơn [33] xác định tỉ lệ từ Hán Việt dùng trong các VB tiếng Việt thuộc các tờ báo phổ biến cho nhiều độc giả sử dụng số lượng từ Hán Việt ở mức trung bình có tỉ lệ nằm ở khoảng 26% khi so với danh sách từ Hán Việt có trong từ điển của Hoàng Phê (từ Hán Việt chiếm khoảng 39,21%) [5]. Tham khảo mức tỉ lệ 26% với kết quả thống kê tỉ lệ sử dụng từ Hán Việt trong truyện ngắn của Lý Lan và Y Ban có thể xác định tới 60 trong tổng 62 truyện có độ khó từ vựng dành cho đối tượng độc giả ở mức trung bình trở xuống.

3.2. Yếu tố trung bình độ dài câu, tổng số từ trong văn bản và tổng số câu trong văn bản

Các nghiên cứu về ĐKVB yếu tố từ Hán Việt trong cấp độ “Từ” luôn được chú trọng. Bên cạnh đó yếu tố trung bình độ dài câu ở cấp độ “Câu” và hai yếu tố “tổng số từ trong văn bản” và “tổng số câu trong văn bản” ở cấp độ “Văn bản” cũng được xem xét và khảo sát để bước đầu phân loại và đánh giá ĐKVB [27-29].

Trong tiếng Việt, theo quan điểm về phương thức từ được cấu tạo thì từ của tiếng Việt có thể xem xét là từ một tiếng, hai tiếng hay nhiều hơn hai tiếng. Nhằm đảm bảo sự phù hợp về mặt lý luận ngôn ngữ liên quan đến đặc điểm loại hình của Tiếng Việt và giữ được sự chính xác, hiệu quả của ngữ liệu được xử lý định lượng từ máy tính, nghiên cứu thực hiện khảo sát bằng cách xem xét yếu tố từ ở phương diện từ một chữ (một tiếng), từ hai chữ (hai tiếng) và từ ba chữ (ba tiếng) qua bộ ngữ liệu gồm 62 truyện ngắn. Bên cạnh đó, độ dài câu được xem là yếu tố bề mặt có ảnh hưởng đến ĐKVB [34], trong đó yếu tố độ dài câu tiếng Việt bao gồm độ dài câu tính bằng kí tự, độ dài câu tính bằng chữ và độ dài câu tính bằng từ. Đề tài thực hiện khảo sát độ dài câu ở số lượng từ với tiêu chí là nhỏ hơn 10 từ và lớn hơn hoặc bằng 10 từ. Đối với số lượng chữ được phân loại theo tiêu chí nhỏ hơn 13 chữ và lớn hơn hoặc bằng 13 chữ [3]. Kết quả khảo sát được trình bày trong Bảng 4.

Bảng 4 Trung bình độ dài câu, tổng số từ và tổng số câu trong truyện ngắn Lý Lan và Y Ban

TT	Tên truyện ngắn	Tổng số từ	Tổng số câu	Trung bình độ dài câu tính bằng từ
1	Biển trong mưa	2.222	288	7,7
2	Tháng chạp	2.405	147	16,4
3	Cuối tuần	1.722	143	12,0
4	Lắp ghép hạnh phúc	1.842	142	13,0
5	Con mèo tưởng đã đi xa	1.600	94	17,0
6	Bay qua bầu trời thành phố đêm giao thừa	1.564	117	13,4
7	Đêm thảo nguyên	2.592	190	13,6
8	Vườn hoàng tử nhỏ	1.410	115	12,3
9	Mẹ và con	2.596	185	14,0
10	Mơ ước hoang đường	1.363	77	17,7
11	Pha lê	1.738	133	13,1
12	Bông vạn thọ	1.499	92	16,3
13	Tâm hồn	1.483	170	8,7
14	Hồi xuân	1.860	156	11,9
15	Romeo và Juliet	1.377	93	14,8
16	Hạnh phúc chơn kinh	1.727	136	12,7
17	Phương pháp hiện thực	1.399	89	15,7

18	Tình chỉ đẹp	1.454	114	12,8
19	Cắm Sơn	1.810	124	14,6
20	Cắm giác	1.060	91	11,6
21	Lòng Hồ	1.602	158	10,1
22	Phi trường Đài Bắc	3.562	250	14,2
23	Xuân thì	2.070	141	14,7
24	Kết thúc có hậu	939	72	13,0
25	Núi không	2.180	176	12,4
26	Tại sao anh làm điều đó	1.493	164	9,1
27	Bé xuân về	1.042	80	13,0
28	Đường dài hạnh phúc	2.502	191	13,1
29	Truyện ma	1.343	141	9,5
30	Một đời	1.008	74	13,6
31	Bạn	904	49	18,4
32	Bức thư gửi mẹ Âu Cơ	6.977	761	9,2
33	Bồ công anh nở bên hồ nước trong	8.493	826	10,3
34	Có thể có và có thể không	7.823	643	12,2
35	Con yêu tinh	6.629	570	11,6
36	Gà ấp bóng	5.003	478	10,5
37	I am đàn bà	7.237	717	10,1
38	Người đàn bà đứng trước gương	3.009	225	13,4
39	Sự nhầm lẫn bò cái	3.339	264	12,6
40	Trong khu vườn nghệ sỹ	3.311	298	11,1
41	Cắm cù	13.757	1.040	13,2
42	Cái Tý	4.941	430	11,5
43	Anh Quảng	4.324	463	9,3
44	Gái góa là gái góa ơi	11.615	1.186	9,8
45	Chín khúc hát ru của thiên thần	4.092	507	8,1
46	Đôi găng tay da màu nâu	2.655	223	11,9
47	Xích lô	2.739	202	13,6
48	Con gái mang cuộc đời mẹ	5.714	523	10,9
49	Jô	2.938	266	11,0
50	Làng Cò	6.812	569	12,0
51	Chồng tôi	2.029	169	12,0
52	Người đàn bà có ma lực	8.940	944	9,5
53	Quê nội	3.072	340	9,0
54	Bản lý lịch tự thuật	3.285	297	11,1
55	Vùng sáng ký ức	5.599	505	11,1
56	Không đề	3.150	273	11,5
57	Chợ rằm dưới gốc dâu cổ thụ	4.475	442	10,1
58	Con quỷ nhỏ trong tôi	6.879	741	9,3
59	Em vẫn gọi tên anh là nước Nga	3.349	343	9,8
60	Mẹ không thể xin lỗi con	4.370	397	11,0
61	Mỗi người đàn ông chỉ của riêng một người đàn bà	5.034	516	9,8
62	Ơi những chú ngựa bắt kham	1.851	172	10,8

Số liệu thống kê trong Bảng 4 cho thấy: trong 31 truyện ngắn Lý Lan (từ thứ tự 1 đến 31) có tổng cộng 4.192 câu, truyện ít nhất có 49 câu (*Bạn*) và truyện nhiều nhất có 288 câu (*Biển trong mưa*). Có thể thấy, dung lượng 31 truyện tính về số lượng câu khá khiêm tốn (dưới 300 câu), số lượng câu trung bình khoảng 135,23. Trong đó, có 10 truyện dưới 100 câu, có đến 19 truyện trên 100 đến dưới 200 câu và chỉ có 2 truyện ngắn trên 200 câu trong tổng 31 truyện. Từ kết quả thống kê, đặc điểm SL trong truyện ngắn Lý Lan được xác định dao động trong khoảng từ 7,7 từ đến 18,4 từ, trong đó có 4 truyện ngắn có SL dưới 10 từ, nằm trong khoảng 7,7 đến 9,5 từ là *Biển trong mưa*, *Tâm hồn*, *Tại sao anh làm điều đó*, *Truyện ma*; có 21 truyện ngắn với SL khoảng 10 từ đến dưới 15 từ nằm trong khoảng từ 10,1 đến 14,8 từ; có 6 truyện ngắn SL trên 15 từ nằm trong khoảng từ 15,7 đến 18,4 từ là *Phương pháp hiện thực*, *Bông vạn thọ*, *Tháng chạp*, *Con mèo tưởng đã đi xa*, *Mơ ước hoang đường*, *Bạn*.

Còn ở truyện ngắn của Y Ban (từ thứ tự 31 đến 62) có tổng 15.330 câu, truyện ít nhất có 169 câu (*Chồng tôi*) và truyện có nhiều câu nhất là 1.186 câu (*Gái góa là gái góa ơi*). Truyện Y Ban có dung lượng khá lớn với số lượng trên 1.000 câu và trung bình số lượng câu là 494,52 trong đó, có đến 17 truyện có số lượng từ trên 100 đến dưới 500 câu; 12 truyện từ trên 500 đến dưới 1.000 câu; còn có 2 truyện trên 1.000 câu trong tổng 31 truyện ngắn. Về đặc điểm SL trong truyện ngắn Y Ban có sự dao động từ 8,1 đến 13,6 từ, trong đó có 9 truyện có SL dưới 10 từ, nằm trong khoảng 8,1 đến 9,8 từ là các truyện: *Chín khúc hát ru của thiên thần*, *Quê nội*, *Bức thư gửi mẹ Âu Cơ*, *Anh Quảng*, *Con quỷ nhỏ trong tôi*, *Người đàn bà có ma lực*, *Gái góa là gái góa ơi*, *Em vẫn gọi tên anh là nước Nga*, *Mỗi người đàn ông chỉ*

của riêng một người đàn bà và 22 truyện ngắn còn lại có SL khoảng 10 từ đến dưới 15 từ nằm trong khoảng từ 10,1 đến 13,6 từ.

Các YTNN trên là những yếu tố được phân tích về mặt định lượng có thể kết hợp các YTNN bề sâu khác tạo nên đặc trưng của thể loại truyện ngắn ở đặc điểm dung lượng. Trong cách biểu hiện phong cách sáng tác, dung lượng cũng góp phần tạo nên đặc trưng trong phong cách của nhà văn. Xem xét tổng thể thì 31 truyện ngắn của Lý Lan có dung lượng rất thấp (dung lượng truyện lớn nhất chỉ có 288 câu) so với dung lượng truyện của Y Ban; cụ thể, chỉ có 8 truyện ngắn trong tổng số 31 truyện ngắn Y Ban có số lượng câu thấp hơn 288, còn lại đều lớn hơn. Ngược lại về SL, ở truyện ngắn Lý Lan lại có đến 11 truyện cao hơn mức 13,6 từ. Kết quả khảo sát đã đưa ra được thống kê số lượng câu làm căn cứ xác định rõ dung lượng của truyện ở mức độ lớn hay nhỏ. Như vậy, không chỉ dung lượng là một yếu tố có ảnh hưởng đến ĐKVB mà còn là các YTNN có thể được xem xét để đánh giá một số biểu hiện ở đặc trưng phong cách sáng tác nhà văn như: thời gian tiếp nhận, giải mã của người đọc và tính hàm súc, sức nén của văn bản khi truyền tải nội dung, nghĩa của từ, ...

3.3 Kết luận về độ khó của truyện ngắn Lý Lan và Y Ban

Các kết quả khảo sát các YTNN được trình bày ở Bảng 1, Bảng 2 và Bảng 3, nghiên cứu cho thấy các YTNN này được xác định qua các con số định lượng cụ thể. Trên cơ sở này, bài viết tiếp tục áp dụng công thức đo ĐKVB: $RL = 0,27WD + 0,13SL + 1,74$ và thang đo đối chiếu để phân loại độ khó của các truyện ngắn (mức độ và cấp lớp) dựa vào chỉ số độ khó [17] đã xác định được cấp độ khó của các truyện ngắn của Lý Lan và Y Ban. Kết quả thu được thể hiện trong Bảng 5.

Bảng 5 Chỉ số độ khó của truyện ngắn Lý Lan

TT	Tên Truyện ngắn	Chỉ số độ khó	Mức độ	Cấp lớp
1	Biển trong mưa	8,5	Khó vừa phải	9-10
2	Tháng chạp	9,3	Khó vừa phải	9-10
3	Cuối tuần	8,9	Khó vừa phải	9-10
4	Lắp ghép hạnh phúc	9,0	Khó vừa phải	9-10
5	Con mèo tưởng đã đi xa	8,8	Khó vừa phải	9-10
6	Bay qua bầu trời thành phố đêm giao thừa	9,6	Khó vừa phải	9-10
7	Đêm thảo nguyên	9,3	Khó vừa phải	9-10
8	Vườn hoàng tử nhỏ	8,9	Khó vừa phải	9-10
9	Mẹ và con	9,8	Khó vừa phải	9-10
10	Mơ ước hoang đường	9,0	Khó vừa phải	9-10

11	Pha lê	9,5	Khó vừa phải	9-10
12	Bông vạn thọ	9,2	Khó vừa phải	9-10
13	Tâm hồn	10,6	Khó	11-12
14	Hồi xuân	9,4	Khó vừa phải	9-10
15	Romeo và Juliet	9,3	Khó vừa phải	9-10
16	Hạnh phúc chơn kinh	10,7	Khó	11-12
17	Phương pháp hiện thực	9,2	Khó vừa phải	9-10
18	Tình chỉ đẹp	9,6	Khó vừa phải	9-10
19	Cắm Sơn	8,7	Khó vừa phải	9-10
20	Cảm giác	9,5	Khó vừa phải	9-10
21	Lòng Hồ	8,8	Khó vừa phải	9-10
22	Phi trường Đài Bắc	9,1	Khó vừa phải	9-10
23	Xuân thì	10,0	Khó vừa phải	9-10
24	Kết thúc có hậu	9,8	Khó vừa phải	9-10
25	Núi không	9,5	Khó vừa phải	9-10
26	Tại sao anh làm điều đó	8,6	Khó vừa phải	9-10
27	Bé xuân về	9,0	Khó vừa phải	9-10
28	Đường dài hạnh phúc	9,1	Khó vừa phải	9-10
29	Truyện ma	9,0	Khó vừa phải	9-10
30	Một đời	8,7	Khó vừa phải	9-10
31	Bạn	9,4	Khó vừa phải	9-10

Kết quả từ Bảng 5 số liệu thu được cho thấy, có đến 29 truyện ngắn ở độ khó phù hợp với đối tượng người đọc ở trình độ lớp 9 và lớp 10 (Khó vừa phải); còn lại 2 truyện là *Tâm hồn*, *Hạnh phúc chơn kinh* thuộc độ khó sử dụng cho đối tượng đọc ở trình độ lớp 11 và lớp 12 (Khó).

Bảng 6 Chỉ số độ khó của truyện ngắn Y Ban

TT	Tên Truyện ngắn	Chỉ số độ khó	Mức độ	Cấp lớp
1	Bức thư gửi mẹ Âu Cơ	8,4	Khó vừa phải	9-10
2	Bồ công anh nở bên hồ nước trong	8,5	Khó vừa phải	9-10
3	Có thể có và có thể không	9,2	Khó vừa phải	9-10
4	Con yêu tinh	9,2	Khó vừa phải	9-10
5	Gà ấp bóng	8,4	Khó vừa phải	9-10
6	I am đàn bà	9,7	Khó vừa phải	9-10
7	Người đàn bà đứng trước gương	8,4	Khó vừa phải	9-10
8	Sự nhầm lẫn bò cái	9,1	Khó vừa phải	9-10
9	Trong khu vườn nghệ sỹ	8,1	Tiêu chuẩn	7-8
10	Cắm cù	9,2	Khó vừa phải	9-10
11	Cái Tý	8,5	Khó vừa phải	9-10
12	Anh Quảng	8,2	Khó vừa phải	9-10
13	Gái góa là gái góa ơi	8,3	Khó vừa phải	9-10
14	Chín khúc hát ru của thiên thần	8,5	Khó vừa phải	9-10
15	Đôi găng tay da màu nâu	8,1	Tiêu chuẩn	7-8
16	Xích lô	8,3	Khó vừa phải	9-10
17	Con gái mang cuộc đời mẹ	8,0	Tiêu chuẩn	7-8
18	Jô	8,0	Tiêu chuẩn	7-8

19	Làng Cò	9,0	Khó vừa phải	9-10
20	Chồng tôi	9,0	Khó vừa phải	9-10
21	Người đàn bà có ma lực	8,1	Tiêu chuẩn	7-8
22	Quê nội	8,7	Khó vừa phải	9-10
23	Bản lý lịch tự thuật	9,6	Khó vừa phải	9-10
24	Vùng sáng ký ức	8,7	Khó vừa phải	9-10
25	Không đề	8,8	Khó vừa phải	9-10
26	Chợ rằm dưới gốc dâu cổ thụ	7,6	Tiêu chuẩn	7-8
27	Con quý nhỏ trong tôi	9,0	Khó vừa phải	9-10
28	Em vẫn gọi tên anh là nước Nga	8,5	Khó vừa phải	9-10
29	Mẹ không thể xin lỗi con	8,4	Khó vừa phải	9-10
30	Mỗi người đàn ông chỉ của riêng một người đàn bà	7,9	Tiêu chuẩn	7-8
31	Ồi những chú ngựa bất kham	9,3	Khó vừa phải	9-10

Số liệu thống kê từ Bảng 6 cho thấy khi phân loại truyện ngắn Y Ban: có được 7 truyện ngắn ở mức độ tiêu chuẩn phù hợp với đối tượng người đọc ở trình độ lớp 7 và lớp 8 là: *Trong khu vườn nghệ sỹ, Đôi găng tay da màu nâu, Con gái mang cuộc đời mẹ, Jô, Người đàn bà có ma lực, Chợ rằm dưới gốc dâu cổ thụ, Mỗi người đàn ông chỉ của riêng một người đàn bà*; còn lại 24 truyện thuộc mức độ khó phù hợp sử dụng cho đối tượng đọc ở trình độ lớp 9 và lớp 10.

Qua khảo sát chỉ số độ khó của 62 truyện ngắn ở cả hai tác giả, số lượng truyện ngắn ở mức độ đọc “khó vừa phải” chiếm đa số, 53 trên 62 truyện với 85,5%; có 7 truyện ngắn ở mức “tiêu chuẩn”, chiếm 11,3%; 2 truyện ở mức “khó”, chiếm 3,2%. Cụ thể thì số truyện ở mức “khó vừa phải” ở Lý Lan có tỉ lệ cao hơn là 93,55% so với Y Ban là 77,42%. Như vậy, đối chiếu kết quả thống kê phân loại cấp độ khó theo lớp của các truyện ngắn Y Ban với truyện ngắn Lý Lan cho thấy, truyện ngắn Lý Lan dành cho đối tượng người đọc có trình độ lớp 9, 10, 11 và 12; trong khi đó, các truyện của Y Ban lại dành cho người đọc có trình độ lớp 7, 8, 10.

4 Thảo luận và đề xuất

Nghiên cứu trình bày kết quả khảo sát, định lượng và phân tích các YTNN theo các câu hỏi nghiên cứu. Kết quả nghiên cứu đã xác định được cấp độ khó của các truyện ngắn thông qua 4 YTNN là: (1) từ Hán Việt, (2) trung bình độ dài câu tính bằng từ, (3) tổng số lượng từ trong một văn bản, và (4) tổng số lượng câu trong một văn bản. Kết quả này phù hợp với các nghiên cứu trước về ĐKVB trên thế giới và tại Việt Nam. Ví dụ, tỉ lệ từ Hán Việt được xác định là một trong những yếu tố xuất hiện gây ra độ khó từ vựng trong tiếng Việt, điều này

phù hợp với kết quả các nghiên cứu trước đã xác định được vai trò và mối quan hệ của từ Hán Việt với các YTNN khác khi nó có tác động đến độ chính xác của việc đánh giá ĐKVB [3,26-33]. Bên cạnh đó, trong nghiên cứu công bố 1935 và một công trình công bố được xem điển hình về các YTNN ảnh hưởng đến ĐKVB vào năm 1959, cho thấy, trung bình độ dài câu tính bằng từ là một trong những YTNN quan trọng tác động đến ĐKVB [13,27-29,32,34]. Ngoài ra, trong công thức đo ĐKVB Flesch được sử dụng rộng rãi cũng đã xác định việc đếm số lượng từ, số câu trong văn bản là một trong những YTNN quan trọng để đánh giá ĐKVB [2,3]. Trên cơ sở kết quả phân loại lớp, cấp học theo chỉ số độ khó cho các truyện ngắn của hai tác giả Y Ban và Lý Lan, ta có thể chọn lọc các truyện ngắn đã được phân loại độ khó theo cấp lớp này phục vụ nhu cầu giáo dục thực tiễn. Thêm vào đó, có thể thấy: (1) dựa vào các cấp độ khó đã được xác định trong bài viết, có thể đề xuất việc sử dụng danh sách các truyện ngắn của Lý Lan và Y Ban làm ngữ liệu phục vụ cho quá trình giảng dạy các lớp, cấp học; (2) ngữ liệu về truyện ngắn góp phần bổ sung tính đa dạng vào kho ngữ liệu tiếng Việt, phục vụ cho việc đo và xác định ĐKVB; (3) kết quả là bước đầu khảo sát từ góc nhìn định lượng về các biểu hiện đặc trưng trong phong cách tác giả từ khung lý thuyết ĐKVB. Nghiên cứu đã đạt được những tiêu chí đề ra ban đầu tuy nhiên vẫn còn giới hạn về mặt số lượng vì chỉ nghiên cứu truyện ngắn của hai tác giả Lý Lan và Y Ban với số lượng ngữ liệu là 62 truyện ngắn. Về mặt phân tích ngữ liệu, chỉ dừng lại ở khảo sát các YTNN đặc trưng bề mặt, có gợi mở hướng đi vào các YTNN ở đặc trưng bề sâu và phân tích biểu hiện đặc trưng của phong cách tác giả. Kết quả bài viết

là cơ sở cho các nghiên cứu mở rộng và sâu hơn ở các YTNK khác đối với ngữ liệu truyện ngắn của Lý Lan

và Y Ban nói riêng và truyện ngắn Việt Nam của các tác giả khác nói chung.

Tài liệu tham khảo

1. Gray, W. S., & Leary, B. E. (1935). *What makes a book readable*. Chicago, Illinois: The University of Chicago Press.
2. Flesch, R. (1948). Teaching bureaucrats plain English. *College English*, 7(8), 470-474.
3. Điệp, N. T. & Vinh, A. L. (2017). Khảo sát các yếu tố ngôn ngữ ảnh hưởng đến độ khó của văn bản – ngữ liệu TOEIC, cấp độ cơ bản. *Tạp chí Khoa học và Đào tạo, Trường Đại học Công nghệ Sài Gòn*, 97-107.
4. Thủy, N. T. T., & Điệp, N. N. (2022). Trích xuất danh mục khóa cạnh sử dụng BERT với hàm mất cân bằng: Aray. *Journal of Science and Technology on Information and Communications*, 1(3), 31-39.
5. Điệp, N. T. (2021). *Các yếu tố ngôn ngữ ảnh hưởng đến độ khó của văn bản tiếng Việt: Đối chiếu với tiếng Anh*. Luận án Tiến sĩ, Trường Đại học Khoa học Xã hội và Nhân văn Thành phố Hồ Chí Minh.
6. Phụng, H.K. (2009). *Đặc trưng truyện ngắn Lý Lan* (Luận văn Thạc sĩ Trường Đại học Sư phạm Thành phố Hồ Chí Minh)
7. Thảo, V.P. (2009). *Đặc điểm văn xuôi Y Ban* (Luận văn Thạc sĩ, Đại học Quốc gia Hà Nội).
8. Hà, T.T. (2011). *Vấn đề nữ quyền trong sáng tác của Y Ban* (Luận văn Thạc sĩ, Trường Đại học Sư phạm, Đại học Thái Nguyên).
9. Trang, D.T. (2012). *Thế giới nghệ thuật trong truyện ngắn Y Ban* (Luận văn Thạc sĩ, Trường Đại học Khoa học Xã hội và Nhân văn, Hà Nội).
10. Ân, L.N. (2004). *Từ điển văn học (bộ mới)*. Nhà xuất bản Thế giới.
11. Giáp, N. T. (2010). *777 khái niệm ngôn ngữ học*. Nhà xuất bản Đại học Quốc gia Hà Nội.
12. Lively, B. A., & Pressey, S. L. (1923). A method for measuring the vocabulary burden of textbooks. *Educational Administration and Supervision*, 9(7), 389-398.
13. Gray, W. S., & Leary, B. E. (1935). *What makes a book readable*. Chicago, Illinois: The University of Chicago Press.
14. Mc Laughlin, G. H. (1969). SMOG grading-a new readability formula. *Journal of Reading*, 12(8), 639-646.
15. Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221-233.
16. Graesser, A. C. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193-202.
17. Nguyen, L. T., & Henkin, A. B. (1985). A second-generation readability formula for Vietnamese. *Journal of Reading*, 29(3), 219-225.
18. Hồng, T.T, & các cộng sự. (2017). *Áp dụng độ khó của văn bản trong việc xây dựng ngữ liệu giáo trình tiếng Việt cho người nước ngoài*. TPHCM: NXB Đại học Quốc gia TP HCM.
19. Sơn, N., Hoàng, H. V., Điệp, N. T. N., Điền, Đ., & Vinh, L. A. (2021). So sánh các văn bản tiếng Việt theo độ khó. *Hue University Journal of Science: Social Sciences and Humanities*, 130(6A), 123-130.
20. Nguyễn, Đ. T. N., Lương, A. V., & Điền, Đ. I. N. H. (2019). Affection of the part of speech elements in Vietnamese text readability. *Acta Linguistica Asiatica*, 9(1), 105-118.
21. Edgar, D. and Chall, J.S. (1948). A Formula for Predicting Readability: Instructions. *Educational Research Bulletin*, 27:37-5 Feb 18, 1948
22. Bailin, A., & Grafstein, A. (2016). *Readability: Text and Context*. Palgrave Macmillan UK. <https://doi.org/10.1057/9781137388773>
23. Lan, L. (2009). *Hồi xuân*. Thành phố Hồ Chí Minh: NXB Văn nghệ.
24. Thắng, B.V tuyển chọn, g. t. (2001). *Truyện ngắn 4 cây bút nữ*. Hà Nội: NXB Văn học.
25. Ban, Y. (2020). *Truyện ngắn Y Ban*. Hà Nội: NXB Hội Nhà văn.

26. Trung tâm Ngôn ngữ học Tính toán Trường Đại học Khoa học Tự nhiên – Đại học Quốc gia TP. HCM (www.clc.hcmus.edu.vn).
27. Đinh Điền. (2006). *Xử lý ngôn ngữ tự nhiên*. NXB Đại học Quốc gia TP. Hồ Chí Minh.
28. Brinton, J. E., & Danielson, W. A. (1958). A factor analysis of language elements affecting readability. *Journalism Quarterly*, 35(4), 420-426.
29. Stolorow, L. M., & Newman, J. R. (1959). A factorial analysis of objective features of printed language presumably related to reading difficulty. *The Journal of Educational Research*, 52(7), 243-251.
30. Maspero, H. (1912). Étude sur la phonétique de la langue annamite. *Bull. de*, 1.
31. Lê Đình Khẩn (2001), *Từ vựng gốc Hán trong tiếng Việt*, NXB Đại học Quốc Gia, Hà Nội.
32. Luong, A. V., Nguyen, D., & Dinh, D. (2018). Assessing the Readability of Literary Texts in Vietnamese Textbooks. In *2018 5th NAFOSTED Conference on Information and Computer Science (NICS)* (pp. 231-236). IEEE.
33. Nguồn ngữ liệu trích từ kho ngữ liệu CLC, Trung tâm Ngôn ngữ học Tính toán, Trường Đại học Khoa học Tự nhiên – Đại học Quốc gia TP. HCM
34. Gray, W. S., & Leary, B. E. (1935). *What makes a book readable*. Chicago, Illinois: The University of Chicago Press.

The short stories of Ly Lan and Y Ban from a view of text readability theory

Diep Thi Nhu Nguyen^{1,*}, Khanh Linh Trinh^{2,3,**}

¹Vietnam Aviation Academy, ²Nguyen Tat Thanh University, ³Sai Gon University

*diepntn@vaa.edu.vn; **linhtk@ntt.edu.vn

Abstract Readability or text readability has long been studied worldwide in both the fields of applied linguistics and computational linguistics, in which the research is mainly for English and some other languages with many high application results. In Vietnamese, the research in terms of readability is still “a low language resource”, especially for short story corpus. Using mainly quantitative method, the paper surveys the linguistic factors that affect the difficulty of the short stories Ly Lan and Y Ban from the perspective of textual difficulty theory. Surveying and analyzing a corpus of 62 short stories written by Ly Lan and Y Ban, the research results show that the readability level of Ly Lan and Y Ban's short stories is mostly at the "medium" level. On the result of the research, the paper offers some applicable suggestions to use texts in appropriate grades and levels, update and supplement the Vietnamese language corpus in research on textual difficulty as well as incorporating this theoretical framework into applications teaching practice, preparing textbooks, and determining the author's style characteristics from a quantitative view.

Keywords readability, linguistic factors, short stories, grades and levels, applicable suggestions.