

ỨNG DỤNG ĐỘ KHÓ CỦA VĂN BẢN TRONG VIỆC GIẢNG DẠY NGÔN NGỮ

● Nguyễn Thị Như Diệp^{1,*} ● Trần Thị Phương Lý²

¹ Trường Đại học Công nghệ Sài Gòn

² Trường Đại học Sài Gòn

TÓM TẮT

Độ khó của văn bản, đặc biệt là trong tiếng Anh, đã được nghiên cứu từ cuối thế kỷ thứ 19 với hàng ngàn công trình được công bố cùng với các đề xuất ứng dụng thực tiễn. Hiện nay vấn đề này vẫn được sự quan tâm của các nhà nghiên cứu không chỉ trong tiếng Anh mà còn nhiều ngôn ngữ khác. Tuy nhiên, tại Việt Nam, việc nghiên cứu “Độ khó của văn bản” vẫn chưa được khai thác nhiều trong việc giảng dạy ngôn ngữ nhằm đưa các giải pháp ứng dụng nhằm đáp ứng nhu cầu thực tế. Trên cơ sở thừa kế các nghiên cứu về độ khó của văn bản, bài viết trước hết trình bày tổng quan về độ khó của văn bản và các yếu tố ngôn ngữ ảnh hưởng đến độ khó của văn bản. Từ đó, dưới góc độ ứng dụng độ khó của văn bản trong thực tiễn, bài viết phân tích các yếu tố ngôn ngữ trên 3 cấp độ ngôn ngữ: “Từ”; “Câu”; và “Văn bản” bằng các nghiên cứu điển hình trong tiếng Anh và tiếng Việt như là các điển cứu minh họa; trên cơ sở này, bài viết đề xuất các giải pháp ứng dụng độ khó của văn bản trong việc dạy ngôn ngữ nói chung, tiếng Anh và tiếng Việt nói riêng tại Việt Nam hiện nay.

Từ khóa: độ khó của văn bản, các yếu tố ngôn ngữ, ứng dụng, giảng dạy ngôn ngữ

APPLICATIONS OF READABILITY IN TEACHING LANGUAGES

● Nguyen Thi Nhu Diep ● Tran Thi Phuong Ly

ABSTRACT

Readability or text readability, especially in English, has been studied since the end of the 19th century with hundreds of thousands of published works and a large number of applied findings. Currently, it is an interesting problem and still being continuously studied by researchers, not only for English but also for many other languages. However, in Vietnam, the study of readability is still limited research to find the solutions to meet the social needs. Based on the findings of readability studies, the paper, initially, presents an overview about readability and the linguistic factors having influence on readability. On the view from readability and applications, the paper investigates and analysis the linguistic factors on 3 main levels: Words, Sentences, and Texts by the illustrative case studies in English and Vietnamese. On this basis, the paper suggests some practical applications of readability in teaching languages in general, English, and Vietnamese in particular in Vietnam currently.

Keywords: readability, linguistic factors, applications, teaching languages

1. ĐẶT VẤN ĐỀ

Độ khó của văn bản, đặc biệt là trong tiếng Anh, đã được nghiên cứu từ cuối thế kỷ thứ 19 với hàng trăm ngàn công trình được công bố cùng với các đề xuất ứng dụng thực tiễn. Hiện nay vấn đề này vẫn được sự quan tâm của các nhà nghiên cứu không chỉ trong tiếng Anh mà còn nhiều

* Tác giả liên hệ: TS. Nguyễn Thị Như Diệp; Email: nhudiep2004@gmail.com

(Ngày nhận bài: 21/09/2022; Ngày nhận bản sửa: 21/10/2022; Ngày duyệt đăng: 11/11/2022)

ngôn ngữ khác. Tuy nhiên, tại Việt Nam, việc nghiên cứu “Độ khó của văn bản” vẫn chưa được khai thác nhiều trong việc giảng dạy ngôn ngữ nhằm đưa các giải pháp ứng dụng nhằm đáp ứng nhu cầu thực tế.

Độ khó của văn bản, có thuật ngữ tiếng Anh là “readability”, năm 1949, Dale và Chall [1] đã đưa ra định nghĩa mang tính khái quát về Độ khó của văn bản là tất cả các tương tác của các yếu tố trong một văn bản in cụ thể, có ảnh hưởng đến sự thành công trong việc đọc tài liệu này của một nhóm độc giả. Sự thành công này là mức độ mà họ đọc hiểu nó, đọc nó với tốc độ tối ưu và cảm thấy nó thú vị. Năm 2008, Fukun Xing, Dongyuan Cheng, & Jianzhong Pu [2] đã đưa ra khái niệm cụ thể hơn về Độ khó của văn bản là cách các tài liệu được viết mà chúng có thể được đọc và hiểu một cách dễ dàng. Độ khó của văn bản phụ thuộc vào nhiều yếu tố, bao gồm (a) độ dài trung bình của câu trong một đoạn văn, (b) số lượng từ mới trong một đoạn văn và (c) độ phức tạp ngữ pháp của ngôn ngữ được sử dụng.

Độ đọc hiểu văn bản (Comprehensibility) và Độ khó của văn bản (Readability) có mối liên quan chặt chẽ với nhau, nhưng về bản chất chúng khác nhau. Hai khái niệm này thường bị nhầm lẫn; nhưng thật ra chúng không được xem là coi là đồng nghĩa vì có sự khác biệt đáng kể giữa chúng trong nhiều nghiên cứu về ĐKVB. Wray và Dahlia (2013) [3] phân biệt rằng **Độ khó của văn bản** (ĐKVB) là một đặc tính của nội tại của chính văn bản trong khi đó **Độ đọc hiểu văn bản** (ĐHV) lại là dấu hiệu cho thấy khả năng đọc hiểu văn bản của chính người đọc. Trong khi các đặc trưng của ĐKVB là nội dung, phong cách, thiết kế và tổ chức của văn bản sẽ giúp cho việc đọc văn bản của độc giả trở nên dễ dàng, thì các đặc trưng của ĐHV là kiến thức có sẵn, kỹ năng đọc hiểu, sự say mê, hứng thú, động lực hoặc thậm chí là tình trạng sức khỏe của độc giả giúp cho việc đọc văn bản trở nên dễ dàng như Dubay (2007) [4] đã phân tích. Có cùng quan điểm này, Chiang et al. (2008) [5] và Jones (1997) [6] cũng đi sâu vào phân tích và cho thấy: ĐKVB là một chỉ số chỉ ra mức độ khó của văn bản, chỉ số này mang tính khách quan, định lượng, cố định và không thay đổi và phụ thuộc vào đặc điểm của người đọc. Trong khi đó, ĐHV bị ảnh hưởng bởi các đặc điểm của người đọc, chẳng hạn như kiến thức nền tảng của người đọc, kiến thức có sẵn, sở thích và khả năng đọc hiểu. Như vậy, đối tượng nghiên cứu của **ĐKVB** là **văn bản** và đối tượng nghiên cứu của **ĐHV** là **người đọc** văn bản.

Lịch sử nghiên cứu về ĐKVB trong các ngôn ngữ (Anh, Nhật, Hoa, Đức, Việt...) cũng như các vấn đề liên quan đến việc đánh giá ĐKVB cho thấy: (1) Nghiên cứu ĐKVB bước đầu từ góc độ xác định các yếu tố của văn bản có ảnh hưởng đến ĐKVB; (2) Các phương pháp, công thức, công cụ đo ĐKVB được phát triển dựa trên các yếu tố ngôn ngữ (YTNN); và (3) Các nghiên cứu và công cụ đo ĐKVB, phần lớn là Tiếng Anh, cho thấy việc xác định ĐKVB được đo lường trên các YTNN thuộc 3 cấp độ chính là: Từ - Câu - Văn bản.

Xuất phát từ câu hỏi nghiên cứu: Các YTNN ảnh hưởng đến ĐKVB trong 3 cấp độ ngôn ngữ “Từ”, “Câu”, và “Văn bản” đã được xác định và ứng dụng như thế nào trong thực tiễn? Trong khuôn khổ và giới hạn của nghiên cứu; bài viết này, dưới góc độ ứng dụng độ khó của văn bản trong thực tiễn, chỉ đi sâu phân tích các yếu tố ngôn ngữ trên 3 cấp độ ngôn ngữ: “Từ”, “Câu”, và “Văn bản” bằng các nghiên cứu điển hình trong tiếng Anh và tiếng Việt như là các điển cứu minh họa; trên cơ sở này, bài viết đề xuất các giải pháp ứng dụng độ khó của văn bản trong việc dạy ngôn ngữ nói chung, tiếng Anh và tiếng Việt nói riêng tại Việt Nam hiện nay.

2. PHƯƠNG PHÁP NGHIÊN CỨU VÀ VẬT LIỆU NGHIÊN CỨU

Từ góc nhìn ứng dụng ĐKVB vào việc giảng dạy ngôn ngữ, bài viết sử dụng phương pháp mô tả và phân tích. Các ví dụ minh họa là các kết quả và các ứng dụng của các công trình nghiên cứu đã công bố và được sử dụng như là các điển cứu minh họa để làm rõ tình ứng dụng thực tế của ĐKVB.

Bên cạnh đó, bài viết sử dụng kết quả có được từ phương pháp thống kê và các thuật toán tin, cũng như công cụ hỗ trợ “CLC- Vietnamese Toolkit”¹ để xử lý số liệu bao gồm 371 văn bản từ Sách giáo khoa Tiếng Việt, Sách giáo khoa Ngữ văn cấp Tiểu học (TH), Trung học cơ sở (THCS) và Trung học phổ thông (THPT).

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Sự ảnh hưởng của yếu tố từ đối với ĐKVB và ứng dụng thực tiễn

3.1.1. Các YTNN bề mặt

Việc xác định các YTNN bề mặt luôn được xem xét hàng đầu với những con số cụ thể cho các cấp độ theo lớp (grade) hoặc cấp độ (level) trong các nghiên cứu về ĐKVB và đã được ứng dụng để đánh giá trong việc phân cấp độ khó cho một văn bản cụ thể một cách định lượng. Chẳng hạn, ta xem xét một ví dụ được trích từ trang Web “Newsinlevel” là trang báo cung cấp thông tin theo ĐKVB như là một ví dụ điển hình và thực tiễn. 3 văn bản sau được viết để cung cấp thông tin cùng một chủ đề về “Động đất ở Indonesia cuối năm 2018” nhưng được chia ra ở 3 cấp độ: Cấp độ 1- Dễ; Cấp độ 2- Trung bình; và Cấp độ 3 - Khó.

Dùng công cụ đo ĐKVB Flessh, được tích hợp sẵn trong Microsoft Word ta cho ra kết quả như sau: tổng số từ trong văn bản 1 là 48; YTNN này tăng lên 110 từ ở cấp độ 2 và đạt tới 185 từ ở cấp độ 3. Điều này cho thấy có mối quan hệ tỉ lệ thuận giữa yếu tố ĐKVB với yếu tố tổng số lượng từ trong văn bản đó, ta có thể thấy điều này trong biểu đồ được minh họa trong ví dụ 1 ở trên.

Bên cạnh những công cụ kiểm tra ngữ pháp và chính tả, phần mềm Microsoft Word còn có thể chấm điểm ĐKVB tiếng Anh bằng công thức Flesch Reading Ease và công thức Flesch Kincaid Grade Level. Phần mềm này có mục đích là xác định ĐKVB phù hợp với nhóm đối tượng ở độ đọc hiểu nào (được kiểm tra trên Microsoft Word của PC và Microsoft Word của Mac). Để sử dụng được chức năng này, vào phần Option \Rightarrow Preview. Đánh dấu check vào 2 ô “Check Grammar with Spelling” và “Show Readability Statistics” (với Mac: Word \Rightarrow Preferences \Rightarrow Spelling and Grammar). Sau khi kiểm tra xong “ngữ pháp và chính tả” (Spelling & Grammar) của VB, một hộp thoại pop-up sẽ hiển thị điểm số của văn bản đó.

ĐKVB được ứng dụng trong phần mềm soạn thảo này sử dụng thang điểm đánh giá của công thức đo ĐKVB Flesch Reading Ease dành cho tiếng Anh [7], chấm theo thang điểm 100 dựa trên các YTNN là từ, câu và đoạn. Điểm càng cao có nghĩa là bài viết càng dễ hiểu. Ngoài ra, Microsoft Word còn tích hợp thêm công thức đo ĐKVB Flesch-Kincaid Grade Level, cũng dùng cho VB tiếng Anh, để xác định ĐKVB theo cấp độ lớp. Cụ thể trong hình minh họa trên, kết quả đo văn bản cho thấy tổng số lượng từ (words) trong văn bản là 178 và số lượng ký tự (characters) là 1026 ký tự, kết quả Flesch Reading Ease là 19.7/100 và Flesch-Kincaid Grade Level là lớp 17.0 cho thấy văn bản này được viết ở cấp độ Khó.

Trên nền tảng tính giá trị về việc xem xét các YTNN bề mặt trong phạm vi từ để đánh giá ĐKVB, bài viết cũng xem xét và xác định các YTNN thuộc phạm vi này trong tiếng Việt với bộ ngữ liệu bao gồm 371 văn bản được trích từ sách giáo khoa Tiếng Việt bậc TH, Ngữ văn bậc THCS và THPT [8], ta có thể thấy các yếu tố bề mặt thuộc phạm vi từ được xác định bằng những con số cụ thể trong bảng sau như một ví dụ minh họa cụ thể.

¹ Công cụ thuộc Trung tâm Ngôn ngữ học Tính toán, Trường Đại học Tự nhiên, Đại học Quốc gia Tp HCM. CLC- <http://www.clc.hcmus.edu.vn>,

Bảng 1. Độ dài văn bản tiếng Việt thuộc phạm vi từ theo bậc học

Độ dài VB tiếng Việt thuộc phạm vi từ theo bậc học						
Bậc học	Tính bằng ký tự		Tính bằng chữ		Tính bằng từ	
	Trung bình	Số lượng từ thấp nhất đến cao nhất	Trung bình	Số lượng từ thấp nhất đến cao nhất	Trung bình	Số lượng từ thấp nhất đến cao nhất
TH	1104	827 – 1.396	231	178 – 288	199	158 – 244
THCS	5295	3709 – 8160	1111	784 – 1.710	946	680 – 1447
THPT	8956	4860 – 10761	1830	1006 – 2.179	1477	862 – 1710

Như vậy, dựa vào số liệu thống kê, ta có thể định lượng một cách khái quát các YTNN bề mặt: (1) trung bình số ký tự, (2) trung bình số chữ và (3) trung bình số từ trong VB viết của 3 bậc lần lượt như sau: Bậc TH là 1104; 231; 199. Bậc THCS là 5295; 1111; 946. Bậc THPT là 8956; 1830; 1477. Từ đó, ta có thể xác định các con số cụ thể để phân loại các văn bản tiếng Việt từ cấp độ cơ bản cho đến trung cấp và đến cao cấp theo các bậc học hoặc cụ thể hơn là theo cấp lớp. Tương tự như vậy, con số này cũng có thể được xác định trong các ngôn ngữ khác nhau như tiếng Đức, Hàn, Trung, Nhật,... thông qua số liệu thống kê từ một bộ ngữ liệu được tuyển chọn, làm sạch và xử lý để phục vụ cho mục đích này.

3.1.2. Yếu tố tần suất từ

Trong tiếng Anh, từ những công thức đo ĐKVB theo kiểu truyền thống, cho đến các quy trình tính toán ĐKVB và cả các công cụ tính toán ĐKVB hiện đại đều xem xét yếu tố tần suất sử dụng của từ để xác định từ khó hay dễ, chẳng hạn hai danh sách từ rất phổ biến: *Thorndike's The Teacher's Word Book* (Từ vựng dành cho Giáo viên của Thorndike) (1921) [9] và *the Dale list* (Danh sách từ của Dale) (Dale 1931; Dale & Chall 1948a; Chall and Dale 1995) [10] là một minh chứng.

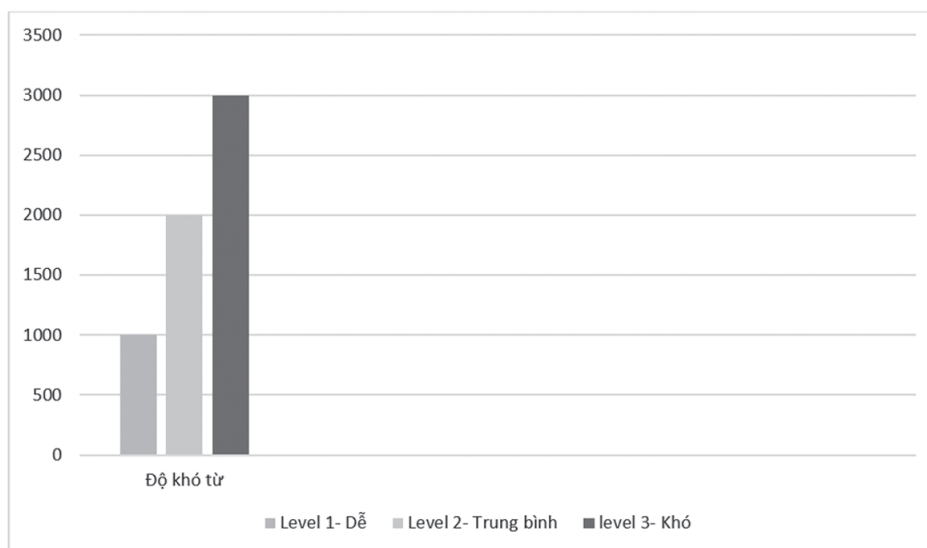
Trong hai danh sách đề cập trên, mỗi danh sách sử dụng một cách tiếp cận khác nhau. Danh sách từ của Thorndike dựa trên các từ được sử dụng trong nhiều nguồn khác nhau và sự phổ biến của từ được xác định với tần suất với các từ xuất hiện trong một kho văn bản. Danh sách từ của Dale (trong phiên bản sửa đổi sau này của ông, Dale & Chall 1948a) bao gồm 3,000 từ được hiểu bởi 80% học sinh lớp bốn, và sự phổ biến của các từ được xác định bởi kiến thức của một nhóm cụ thể đang được thử nghiệm.

Như vậy, chúng ta có thể thấy từ phổ biến được xem xét trong hai danh sách từ trên được tiến hành với hai cách tiếp cận khác nhau: “cách tiếp cận tần số” và “cách tiếp cận tri thức”; do vậy, chúng ta có thể sử dụng một trong hai cách hoặc kết hợp cả hai để trích xuất ra được danh sách từ phổ biến cho từng ngôn ngữ cụ thể và cấp độ cụ thể cho mục đích giảng dạy trong thực tiễn.

Yếu tố tần suất từ luôn là YTNN quan trọng nhất để đánh giá ĐKVB trong phạm vi từ, chẳng hạn, xem xét YTNN này ở ví dụ 3.1 đã được đề cập bên trên, ta thấy: ở cấp độ 1 văn bản sử dụng 1000 từ cơ bản nhất trong Tiếng Anh, từ đơn và từ khó “*earthquake*” cũng được giải thích bằng những từ phổ biến nhất trong 1000 từ cơ bản nhất của tiếng Anh (when the ground moves). Ở cấp độ 2 văn bản được quy định là sử dụng 2000 từ cơ bản nhất trong tiếng Anh; bên cạnh từ đơn thì các từ ghép được sử dụng trong văn bản, từ khó “*magnitude earthquake*” được giải thích và số lượng từ được mở rộng là 3000 từ cơ bản và từ khó “*magnitude earthquake*” không còn được giải thích ở cấp độ 3.

3.2. Sự ảnh hưởng của yếu tố câu đối với ĐKVB và ứng dụng thực tiễn

3.2.1. Các YTNN bề mặt



Hình 1. Trung bình độ dài câu tính bằng từ theo 3 cấp độ khó

Tương tự như trong phạm vi từ, ở mức độ câu các YTNN bề mặt cũng được xác định để đánh giá ĐKVB. Chẳng hạn ta xét văn bản tiếng Anh đã được đo ĐKVB trong hình minh họa 3.1 trên, ta thấy: trung bình số lượng câu trong đoạn văn (Averages: sentences per paragraph) là 6.0, hoặc ta có thể thấy rõ điều này qua số liệu thống kê của các văn bản được phân cấp ở 3 cấp độ: Dễ - Trung bình - Khó ở ví dụ 3.1 đã đề cập trên qua biểu đồ trên Hình 2.

Số liệu thống kê cho thấy YTNN độ dài trung bình câu tính bằng từ ở cấp độ 1 là 6.0, tăng lên 9.1 ở cấp độ 2 và đạt đến 11.5 ở cấp độ 3. Điều này cho thấy việc xác định ĐKVB theo bằng những con số cụ thể là bước cơ bản để định lượng cho độ khó của một văn bản cụ thể.

3.2.2. Yếu tố độ sâu cây cú pháp

Việc định lượng YTNN các yếu tố bề mặt (trung bình: độ dài câu tính theo ký tự, độ dài câu tính bằng chữ, độ dài câu tính bằng từ) trong một văn bản luôn đóng vai trò quan trọng trong bước đầu đánh giá ĐKVB. Tuy nhiên để xem xét và đánh giá mức độ khó của một văn bản nào đó thì việc chỉ xem xét YTNN bề mặt là chưa đủ.

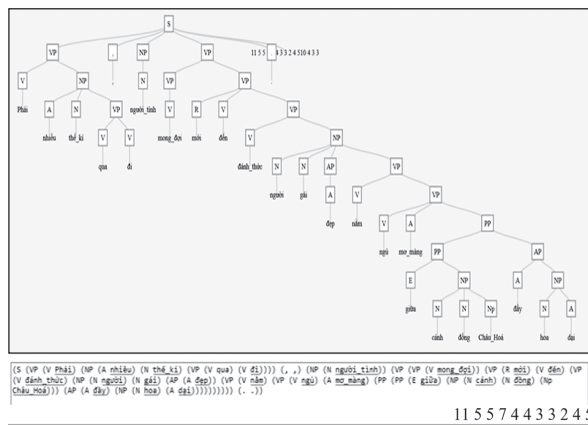
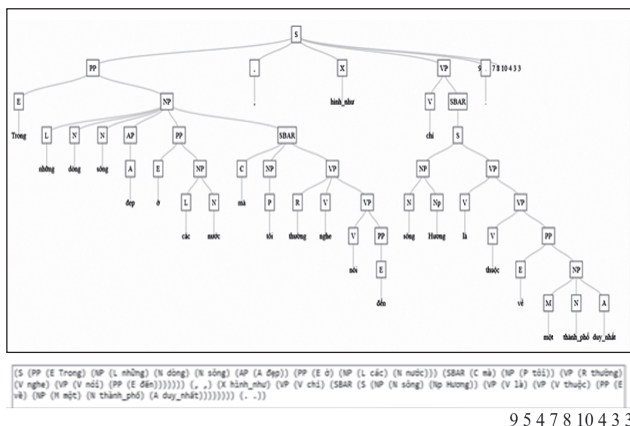
Có thể thấy rằng cấu trúc ngữ pháp càng đơn giản thì câu càng dễ hiểu và ngược lại. Tuy nhiên, việc xác định cú pháp đơn giản hay phức tạp phải được căn cứ trên nhiều YTNN khác nhau như: từ loại, ngữ nghĩa, cấu trúc câu; đặc biệt là các YTNN này có thể được xác định theo cả hướng định tính và định lượng. 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.

Như đã phân tích bên trên, ta thấy việc xác định ĐKVB cần phải kết hợp cả việc đánh giá định lượng lẫn định tính. Do vậy, để có thể định lượng yếu tố ngữ pháp một cách tổng quát và nhằm đánh giá sự ảnh hưởng của yếu tố ngữ pháp đối với ĐKVB cũng như có thể phân loại cú pháp của câu từ đơn giản (cấp độ Dễ) đến phức tạp (cấp độ Khó) đối với một văn bản, ta có thể đánh giá mức độ ảnh hưởng của yếu tố ngữ pháp dựa trên YTNN độ sâu của cây cú pháp [11]. Xét ví dụ sau như một minh họa điển hình.

Ví dụ 3.2:

Văn bản A

Văn bản B



Trước hết, hình biểu diễn trực quan của hai câu A và B cho thấy: (1) Câu A: 23 từ: Trong/ những/ dòng/ sông/ đẹp/ ở/ các/ nước/ mà/ tôi/ thường/ nghe/ nói/ đến/, hình như/ chỉ/ sông Hương/ là/ thuộc/ về/ một/ thành phố/ duy nhất, Câu B: 22 từ: Phải /nhiều/ thế kỉ/ qua/ đi/, người tình/ mong đợi/ mới/ đến/ đánh thức/ người/ gái/ đẹp/ nằm/ ngủ/ mơ màng/ giữa/ cánh đồng/ Châu Hoá/ đầy/ hoa/ dại; (2) trong khi biểu đồ hình cây câu A có các nhánh trải về cả hai bên phải và bên trái, thì biểu đồ hình cây của câu B có xu hướng chỉ nghiêng về phía bên phải; (3) Kết quả cho thấy, độ sâu của cây cú pháp câu A là 9 tầng và độ sâu của cây cú pháp câu B là 11 tầng.

Ta có thể kết luận rằng cây cú pháp A có độ sâu hơn B, do vậy câu B có mức độ cú pháp phức tạp hơn và khó hiểu hơn câu A.

Như vậy, độ sâu của cây cú pháp là YTNN mang tính tổng quát có thể xác định được sự phức tạp về mặt ngữ pháp của văn bản một cách nhanh chóng và hiệu quả. Hay nói cách khác xét về phạm vi ngữ pháp, độ sâu của cây cú pháp là YTNN bề sâu có ảnh hưởng rất lớn để đánh giá ĐKVB tiếng Việt nói riêng và các ngôn ngữ khác nói chung.

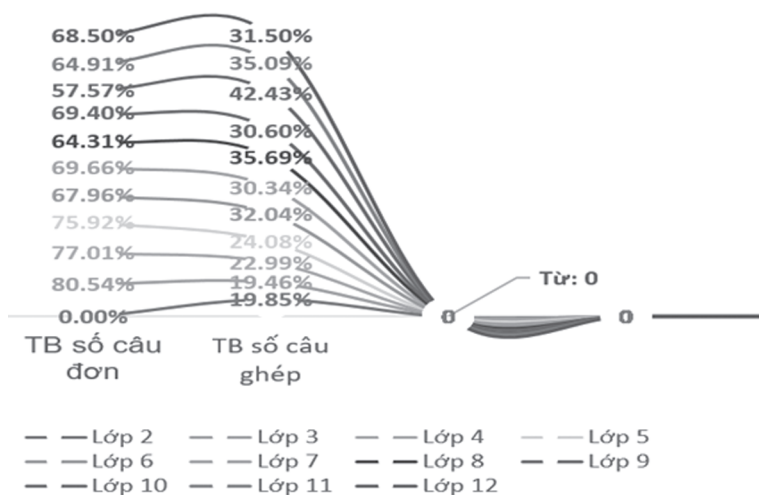
3.3. Sự ảnh hưởng của yếu tố văn bản đối với ĐKVB và ứng dụng thực tiễn

3.3.1. Các YTNN bề mặt

Các nghiên cứu cho thấy ĐKVB không chỉ được đánh giá dựa vào sự ảnh hưởng của các yếu tố từ và yếu tố câu đối với VB, mà còn được đánh giá dựa vào sự ảnh hưởng của các YTNN trong một chỉnh thể của VB. Nghiên cứu về ĐKVB cho thấy yếu tố số lượng câu trong VB luôn được xem xét như là một trong những YTNN bề mặt quan trọng để đánh giá ĐKVB. Chẳng hạn như Vogel và Washburne (1928) [12] đưa YTNN là số lượng câu đơn trong một mẫu VB có 75 câu trong công thức đo ĐKVB; Gray và Leary (1935) [13] xem xét số lượng các loại câu để đánh giá ĐKVB; Lorge (1944) [14] và Flesch (1948) đánh giá ĐKVB gồm nhiều yếu tố trong đó có số lượng câu được sử dụng trong văn bản. Trong tiếng Việt [14], chúng tôi cũng đã khảo sát các YTNN này, ta xét số liệu thống kê trong biểu đồ dưới đây như một minh chứng:

Kết quả thống kê yếu tố số lượng câu đơn và số lượng câu ghép được sử dụng trong một văn bản theo cấp lớp cho thấy đa số chúng có chiều hướng tỉ lệ thuận đối với ĐKVB theo cấp lớp, nghĩa là ở cấp lớp thấp có khuynh hướng dùng câu đơn, và tỉ lệ câu ghép tăng dần khi cấp lớp cao hơn, chẳng hạn lớp 2 tỉ lệ câu đơn dùng ở lớp 2 là hơn 80% và giảm xuống khoảng 75% ở lớp 5 và tiếp tục giảm còn gần 70% ở lớp 6. Như vậy, số lượng câu ghép sẽ tăng lên khi cấp lớp cao hơn, lớp 2 tỉ lệ câu ghép là gần 20% tăng lên khoảng 25% ở lớp 5 và tăng đến hơn 30% ở lớp 6,... Hay nói cách khác,

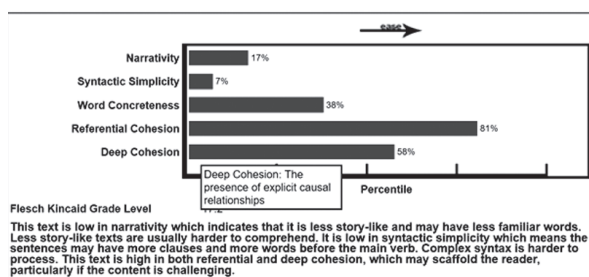
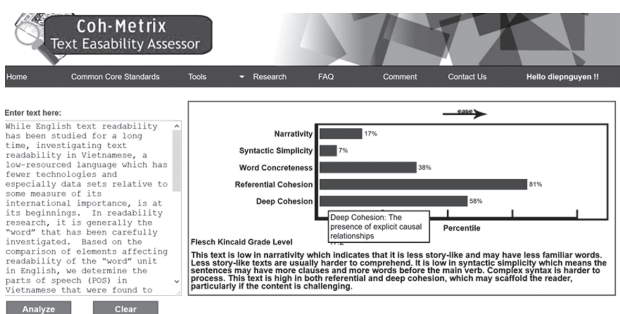
số lượng câu ghép được dùng tăng dần khi ĐKVB tăng dần. Văn bản có nhiều câu ghép, độ sâu của cú pháp của các câu trong văn bản tăng, thì độ phức tạp của kết cấu văn bản tăng và độ khó của văn bản cũng tăng theo.



Hình 2. Trung bình số lượng câu đơn, câu ghép theo cấp độ lớp, ngữ liệu Tiếng Việt-Ngữ Văn

3.3.2. Yếu tố liên kết trong văn bản

Việc xác định YTNN ảnh hưởng đến ĐKVB về mặt kết cấu văn bản được xem xét không chỉ về YTNN 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 VB. Tính liên kết chính là nhân tố quan trọng nhất có tác dụng biến một chuỗi câu thành một văn bản. Chính vì vậy các YTNN chỉ ra các mối liên kết trong văn bản luôn được chú trọng trong các nghiên cứu về ĐKVB. Ta xét một văn bản tiếng Anh đã được đo tự động bằng mô hình đo ĐKVB Coh-Metrix³ như là một ví dụ minh họa minh chứng cho việc ứng dụng điển hình của ĐKVB trong thực tiễn.



Hình 3. Đo ĐKVB tiếng Anh bằng công cụ Coh- Metrix

Sử dụng VB tiếng Anh được trích trong bài viết: Affection of the Part of Speech Elements in Vietnamese Text Readability [15] để đo ĐKVB bằng công cụ Coh- Metrix, kết quả đo độ khó được xác định tỉ lệ phần trăm dựa trên 5 YTNN cụ thể như sau: (1) Narrativity (Sự tường thuật): 17%; (2): Syntactic Simplicity (Sự đơn giản trong cú pháp): 7%; (3) Word Concreteness (Tính cụ thể của từ): 38%; (4) Referential Cohesion (Tính liên kết về sở chỉ): 81%; và (5) Deep Cohesion (Tính liên kết về bề sâu): 58%.

Trên cơ sở xác định 5 YTNN trên, văn bản được đo cho ra kết quả cụ thể như sau: “Văn bản này

³ <http://www.cohmetrix.com>

có chỉ số tường thuật thấp, điều này có nghĩa là vấn đề trình bày không quen thuộc và từ ít phổ biến. Văn bản ít quen thuộc thường khó để đọc hiểu. Độ đơn giản trong cú pháp thấp nghĩa là câu có nhiều thành phần chêm xen và từ đứng trước động từ chính. Cú pháp phức tạp dẫn đến khó xử lý văn bản hơn. Văn bản này có chỉ số cao ở cả hai YTNN liên kết là sở chỉ và liên kết về bề sâu, điều này có thể cung cấp các mối liên kết cho người đọc, đặc biệt là nếu nội dung quá khó”

Đánh giá cấp độ khó của văn bản tiếng Anh được đo, xác định ĐKVB này theo thang đo Flesch Kincaid Grade Level là ở cấp lớp 17.2. Như vậy, rõ ràng với cấp độ lớp là 17.2 theo thang đo Flesch Kincaid Grade Level được xác định bởi công cụ Coh - Metrix cho thấy văn bản trên là rất khó, điều này được thể hiện rất rõ qua 2 YTNN cơ bản nhất là từ và cấu trúc câu với tỉ lệ và con số cụ thể và nhận xét tổng quát được nêu trên. Tuy nhiên, văn bản này được xác định có chỉ số liên kết cao được thể hiện hiển lộ bằng các kết từ, phép thế, phép nối, phép lặp trong VB, từ đó giúp VB dễ đọc hiểu hơn.

Tóm lại, kết quả trên cho thấy việc xác định yếu tố liên kết trong văn bản, chẳng hạn: phép nối, phép lặp, phép thế,... đóng vai trò rất quan trọng trong việc xem xét tổng hòa các mối quan hệ cũng như sự ảnh hưởng của toàn bộ các YTNN cả bề mặt và bề sâu trong văn bản để tiến hành đánh giá và phân loại độ khó cho văn bản.

4. KẾT LUẬN

Nghiên cứu về ĐKVB cho thấy các YTNN ảnh hưởng đến văn bản luôn được chú trọng nghiên cứu, xem xét, phân tích và đánh giá theo cả hướng định tính và định lượng ở cả các YTNN bề mặt và bề sâu trên 3 cấp độ: Từ - Câu - Văn bản.

Trong khuôn khổ của bài viết, không ít vấn đề liên quan đến ĐKVB chỉ được trình bày ở mức nhận diện, miêu tả có tính chất đặt vấn đề, phần lớn theo hướng định lượng chứ chưa có điều kiện trình bày sâu hơn. Mục đích chính của bài viết là làm rõ vấn đề nghiên cứu ĐKVB là gì, xác định vai trò và mức độ ảnh hưởng của các YTNN đến ĐKVB trong 3 cấp độ ngôn ngữ “Từ”, “Câu”, và “Văn bản” bằng các minh chứng và các ứng dụng trong thực tiễn. Trên cơ sở đi vào phân tích các YTNN bề mặt và các YTNN bề sâu đã được chứng minh là có mức độ ảnh hưởng cao trong các cấp độ trên, bài viết đưa ra một số đề xuất ứng dụng ĐKVB trong quá trình giảng dạy ngôn ngữ.

Hiện nay, nhu cầu cải cách hay chuẩn hóa sách giáo khoa, giáo trình cho các cấp đào tạo từ bậc mầm non cho đến bậc đại học ở Việt Nam đang là một vấn đề được Nhà nước và xã hội quan tâm và việc xác định các YTNN có ảnh hưởng đến ĐKVB có thể được ứng dụng để biên soạn tài liệu giảng dạy ngoại ngữ, chẳng hạn chuyển dịch danh sách 1000 từ, 2000 từ, 3000 từ cơ bản của tiếng Việt sang nhiều thứ tiếng khác nhau như: Hàn, Hoa, Nhật, Đức... để đáp ứng nhu cầu giảng dạy thực tế tại Việt Nam. Các danh sách tần suất từ tiếng Việt có thể trợ giúp trong việc giảng dạy tiếng Việt theo khung 6 bậc, xây dựng ngân hàng đề thi tiếng Việt và một số ngoại ngữ khác khi được chuyển dịch.

Việc xác định các YTNN ảnh hưởng đến ĐKVB với những con số cụ thể cũng có thể phục vụ cho nhu cầu thiết thực trong quá trình giảng dạy tiếng Việt hoặc định hướng tương tự cho các ngôn ngữ khác trong giảng dạy ngôn ngữ, chẳng hạn như chọn lọc văn bản, viết văn bản mẫu, biên soạn từ điển, sách hướng dẫn, hợp đồng, tuyển phỏng viên...

Trong giới hạn của bài viết và vì phần lớn bài viết thực hiện theo hướng định lượng nên chưa thể đi sâu phân tích vào từng yếu tố, đặc biệt là các YNNN mang tính định tính, hoặc cả định lượng và định tính có ảnh hưởng đến ĐKVB. Kết quả của bài viết hy vọng sẽ giúp được việc xác định các YTNN ảnh hưởng đến ĐKVB bằng những con số cụ thể trên ba phương diện: Từ – Câu – VB, để chúng ta có thể bước đầu phân loại và xác định được ĐKVB ở một cấp độ cụ thể với số lượng văn bản lớn, nhanh, hiệu quả, tiết kiệm thời gian, công sức và tài chính nhằm đáp ứng nhu cầu giảng dạy ngoại ngữ.

TÀI LIỆU THAM KHẢO

- [1] Dale, E., & Chall, J. S., “The Concept of Readability. Elementary English,” 26(1), pp 19-26, 1949.
- [2] Fukun Xing, Dongyuan Cheng & Jianzhong Pu, “A New Approach to Readability Study Based on Information Computing,” Proceedings - ALPIT 2008, 7th International Conference on Advanced Language Processing and Web Information Technology, pp. 156–161, 2008.
- [3] Wray, D. & J. Dahlia., “Exploring the Readability of Assessment Tasks: The Influence of Text and Reader Factors,” REMIE: Multidisciplinary Journal of Educational Research, 3(1), p. 73, 2013.
- [4] Dubay, H. W., “Smart Language: Readers, Readability, and the Grading of Text. Impact Information,” Costa Mesa: California, pp. 4-6, 2007.
- [5] Chiang, W.-C., T.D. Englebrecht, T.J. Phillips Jr & Y. Wang., “Readability of financial accounting principles textbooks,” The Accounting Educators’ Journal, 18, p. 48, 2008.
- [6] Jones, M.J., “Methodological themes: Critical appraisal of the Cloze procedure’s use in the accounting domain,” Accounting, Auditing and Accountability Journal, 10(1), pp. 105–106, 1997.
- [7] Flesch, R. F. A., “New Readability Yardstick,” Journal of Applied Psychology, 32(3), pp. 221-233, 1948.
- [8] An-Vinh Luong, Diep Nguyen, Dien Dinh, “Assessing the Readability of Literary Texts in Vietnamese Textbooks”, 2018 5th NAFOSTED Conference on Information and Computer Science (NICS). HoChiMinh city, Vietnam, pp. 233-238, 2018.
- [9] Thorndike, E. L., “The Teacher’s Word Book,” New York: Teachers College, Columbia University, 1921.
- [10] Dale, E., & Chall, J. S. (1949). “The Concept of Readability,” Elementary English, 26(1), pp. 19-26, 1949.
- [11] Nguyễn Thị Như Diệp, “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, Tp. HCM, 2021.
- [12] Vogel, M., & C. Washburne., “An Objective Method of Determining Grade Placement of Children’s Reading Material,” The Elementary School Journal, 28(5), pp. 373– 381, 1928.
- [13] Gray, S. W., & Leary, E. B, “What Makes a Book Readable,” Chicago, Illinois: The University of Chicago Press, 1935.
- [14] Lorge, I., “Predicting Readability,” Teachers College Record, 45(6), pp. 404– 419, 1944.
- [15] Diệp Thị Nhu Nguyễn, An-Vinh Luong, Điền Đình, “Affection of the part of speech elements in Vietnamese text readability,” Acta Linguistica Asiatica, 9(1). doi:10.4312/ala.9.1.105-118, 2019.