

Nhân ngày Tết:

ĐỌC THƠ LỤC BÁT BẰNG... THỐNG KÊ

Trong các thể thơ Việt Nam, lục bát có lẽ là một trong những thể loại thông dụng nhất. Trong số 168 bài thơ do Hoài Thanh và Hoài Chân tuyển chọn và phê bình trong *Thi Nhân Việt Nam*, có đến 26 bài (tức 15%) được viết theo cấu trúc sáu-tám. *Truyện Kiều* của Nguyễn Du, và *Lục Vân Tiên* của Nguyễn Đình Chiểu là hai tác phẩm tiêu biểu cho thể thơ này. Trong phong trào *Thơ Mới*, có lẽ Nguyễn Bính là một người viết tương đối nhiều thơ lục bát hơn các nhà thơ nổi tiếng khác như Xuân Diệu, Thế Lữ, Huy Cận, Chế Lan Viên v.v...

Có lẽ qua *Kiều*, mà thể thơ lục bát trở nên rất phổ thông trong người dân Việt. Hầu như mọi người Việt Nam, từ thành phần hàn lâm tới dân lao động, thậm chí người không đọc được chữ, đều có thể thuộc vài câu thơ lục bát, hoặc từ *Truyện Kiều*, hoặc từ *Lục Vân Tiên*, hoặc thuộc lấy đôi ba câu thơ của Nguyễn Bính và lấy làm tâm đắc. Nhiều câu văn, nhiều chữ trong *Truyện Kiều* đã đi vào kho tàng ca dao tục ngữ.

Đã từ lâu, tôi vẫn thường tự hỏi tại sao một tác phẩm văn chương được viết bằng thể thơ sáu-tám, được lặp đi lặp lại cả ngàn lần và có thể nói là ... nhàm chán, nhưng lại được đại đa số quần chúng ngâm đi ngâm lại không biết chán. Thậm chí, thể thơ lục bát còn được các nhà thơ sau này mô phỏng theo, và cũng được nhiều người đọc ủng hộ. Có phải tại người Việt chúng ta quá dễ dãi với thơ văn, thích theo đường mòn, thiếu tính khai phá, hay tại vì thể thơ lục bát quá hay, quá hợp với ngôn ngữ Việt và tâm tình người Việt? Trong *Truyện Kiều* có những "công thức" thơ gì và sự phân phối từ ngữ như thế nào để có thể thu hút độc giả một cách mạnh mẽ như thế?

Để có khả năng hấp dẫn người đọc như thế, thể thơ lục bát chắc hẳn phải có một sức thu hút lớn. Trong vòng một trăm năm qua, đã có rất nhiều công trình nghiên cứu về *Truyện Kiều* dựa trên

nhiều góc độ và phương pháp khác nhau. Trong hàng trăm nghiên cứu đó, một số được tập trung vào ngôn ngữ dùng trong *Truyện Kiều*, và đã khám phá ra nhiều điều khá thú vị và sâu sắc. Tuy nhiên, những công trình nghiên cứu về ngôn ngữ trong quá khứ mang nặng tính cách nghiên cứu văn chương và văn học; tức là, người nghiên cứu nhằm vào việc tìm ra các mối quan hệ giữa chữ nghĩa và tư tưởng hay các giá trị đạo lý và xã hội của tác phẩm. Nhưng hình như chưa có nghiên cứu nào đặt *Truyện Kiều* hay thể thơ lục bát dưới lăng kính của khoa học định lượng, như toán học chẳng hạn. Nói một cách khác, những nghiên cứu trong quá khứ chỉ tập trung vào định chất (qualitative research), chứ không định lượng (quantitative research).

Mà, văn học và khoa học có một mối tương giao thâm kín. Có người cho rằng (và tôi đồng ý) văn học và khoa học không hẳn là hai bộ phận tách biệt. Trong nghệ thuật có khoa học, và trong khoa học có nghệ thuật. Nguyên tắc âm luật trong thơ, nhạc, v.v... có thể coi như là một phát biểu khoa học của nghệ thuật. Tỷ lệ vàng có thể coi như một đặc tính nghệ thuật trong khoa học (kiến trúc). Nếu chỉ biết nguyên tắc âm luật không (chỉ biết khoa học) thì chưa chắc người nhạc sĩ soạn được một bản nhạc hay. Nếu chỉ biết "vẻ đẹp" không (nghệ thuật) chưa chắc nhà kiến trúc có thể xây một lâu đài đẹp. Do đó, một giả thuyết được đặt ra là trong thơ lục bát có một cấu trúc mượt mà có khả năng thu hút người đọc.

Nhằm mục đích khai thác các câu hỏi và giả thuyết nêu trên, tôi mạo muội làm một vài phân tích thống kê về qui luật trong thơ lục bát. Tôi chọn ba tác phẩm chính mà tôi nghĩ là tiêu biểu cho thể thơ lục bát để phân tích. Đó là *Truyện Kiều* (sẽ gọi tắt là *Kiều* trong bài viết này) của Nguyễn Du, *Lục Vân Tiên* của Nguyễn Đình Chiểu, và *Lữ bước sang ngang* của Nguyễn Bính. Ba tác phẩm được sáng

tác vào ba thời kì khác nhau trong lịch sử văn học Việt Nam, và sự khác biệt về thời gian có thể phản ánh vài thay đổi về cấu trúc từ ngữ thơ lục bát, tức là một điều cần nghiên cứu. Lí do nữa là ba tác phẩm này tương đối dài (hơn 500 câu thơ) so với những bài thơ ngắn, và độ dài của các tác phẩm này giúp cho việc ước tính [bằng phương pháp thống kê học] có độ chính xác và độ tin cậy cao.

Phân phối từ ngữ

Trong *Kiểu*, với 3.254 câu thơ, Nguyễn Du dùng 22.778 từ; trong số đó, chỉ có 2.407 là từ gốc [1]. Nói cách khác, tính trung bình, số lần lặp lại cho mỗi từ là 9,5 lần. Trong *Lục Vân Tiên*, Nguyễn Đình Chiểu dùng 14.518 từ, nhưng chỉ 2.533 từ duy nhất, và tính trung bình số lần lặp lại cho mỗi từ là 5,7 lần. Trong bài *Lữ bước sang ngang*, Nguyễn Bính dùng 1.862 từ, trong đó có 605 từ duy nhất, và độ lặp lại trung bình là 3,1 lần cho mỗi từ. Sự khác biệt này cũng có thể đoán được, vì mức độ sử dụng từ thông thường tăng theo tỉ lệ thuận với độ dài của một tác phẩm.

Độ dài của chữ được tính bằng cách đếm số mẫu tự [2] trong chữ đó. Dùng số từ-duy-nhất làm mẫu số, tính trung bình, mỗi từ đơn trong *Kiểu* dài khoảng 3,45 mẫu tự với độ lệch tiêu chuẩn (standard deviation) [3] là 0,93 mẫu tự, và hệ số biến thiên khoảng 27%. Những chỉ số thống kê này rất tương đương với độ dài của chữ trong *Lục Vân Tiên* và *Lữ bước sang ngang*. Khoảng 74% trong tổng số từ-duy-nhất có độ dài từ 3 đến 4 mẫu tự, và chỉ có khoảng 14% dài từ 1 đến 2 mẫu tự. Chữ dài nhất trong *Kiểu* có 7 mẫu tự, đó là chữ *Nghiêng* được dùng tất cả là 5 lần.

Theo các nhà nghiên cứu ngôn ngữ học, nguyên âm được chia ra làm hai loại trầm và bổng tùy theo độ khép hay mở của môi khi phát âm. Những **nguyên âm** này có thể tóm tắt như sau: âm bổng (*i, ê, e*), trung (*ư, ơ, â, a, ă*) và trầm (*u, ô, o*). **Phụ âm** cuối trong tiếng Việt cũng có thể chia thành hai nhóm: vang và tắc. Phụ âm vang gồm có *m, n, nh, ng*; và phụ âm tắc gồm: *p, t, ch* và *c*.

Dùng cách phân loại này, trong *Kiểu* có khoảng 21% từ nguyên âm bổng, tương đương với *Lữ bước sang ngang* (21%), nhưng cao hơn so với *Lục Vân*

Tiên (19,5%). Ngược lại, trong thơ Nguyễn Bính dùng từ có nguyên âm trầm lên đến 35%, cao hơn so với Nguyễn Du (31%) và Nguyễn Đình Chiểu (31%). Trong thơ lục bát, Nguyễn Bính dùng từ có phụ âm vang lên đến 81,4%, cao hơn Nguyễn Du (77%) và Nguyễn Đình Chiểu (80%).

Phân tích những kết hợp giữa phụ âm và nguyên âm cho thấy phần lớn những phụ âm vang là vang-trung (40% trong *Kiểu*, 43% trong *Lục Vân Tiên*, và 38% trong *Lữ bước sang ngang*), tiếp theo là vang-trầm (23% trong *Kiểu*, 25% trong *Lục Vân Tiên*, và 28% trong *Lữ bước sang ngang*). Các từ có phụ âm tắc lại tập trung vào tắc-trung, và tắc-trầm; trong khi đó phụ âm tắc-bổng chỉ chiếm khoảng 1,2% đến 1,8%. (Có thể tham khảo thêm chi tiết trong Bảng 1).

Phân phối vần điệu

Nếu tính theo qui tắc “chuẩn” của thơ lục bát (*bb tt bb / bb tt bb tb*), một bài thơ lục bát dài phải có 64% chữ thanh bằng và 36% chữ thanh trắc. Nhưng trong thơ lục bát, cũng như nhiều thể thơ khác, có một vài ngoại lệ, và do đó sự phân phối thanh bằng trắc không nhất thiết phải theo tỉ lệ trên. Những ngoại lệ này là các chữ số lẻ (1, 3, 5 và 7) trong cả hai câu sáu và tám chữ có thể là thanh bằng hay thanh trắc. Do đó trong thực tế, số lượng từ thanh bằng ít hơn, và từ thanh trắc nhiều hơn, công thức trên. Trong số 22.778 từ trong *Kiểu*, có gần 60% là từ thanh bằng, và 40% từ thanh trắc. Độ phân phối này cũng rất tương đương với độ phân phối trong *Lục Vân Tiên* (61% thanh bằng và 39% thanh trắc). Tuy nhiên, trong thơ của Nguyễn Bính có khuynh hướng thanh bằng nhiều hơn (64%) so với Nguyễn Du và Nguyễn Đình Chiểu.

Dựa vào ngoại lệ nằm trong khuôn khổ được cho phép trên, luật bằng trắc có thể đơn giản hóa thành *bttb* và *btbb* cho những chữ ở vị trí số 2, 4, 6 (cho câu sáu chữ) và 2, 4, 6, 8 (cho câu tám chữ). Theo “công thức đơn giản hóa” này, nhà thơ “được phép” chọn tất cả 8 biến thể (variations) khả dĩ cho câu sáu chữ, và 16 biến thể khả dĩ cho câu tám chữ. Tuy nhiên, đối với câu sáu chữ, Nguyễn Du không những dùng tất cả 8 biến thể khả dĩ cho phép, mà còn sáng tạo thêm 14 biến thể “mới” với

29 câu thơ. Trong số 29 câu “phá luật” này, 19 câu tập trung ở công thức *ttb*, 6 câu theo *bbb*, và 4 câu theo *tbb*. Trong số 29 câu phá lệ này, có đến 26 câu được viết theo nhịp 3/3, như “*khi hương sớm, khi trà trưa*” hay “*Nền phú hậu, bậc tài danh*”, và chỉ có ba câu không theo nhịp trên (*sao chẳng biết ý tứ gì, tưởng bây giờ là bao giờ, mụ quản gia vãi Giấc duyên*). Điều thú vị là trong những câu 3/3 này, chữ “*kh*” thường được dùng đi dùng lại khá nhiều lần. Trong khi Nguyễn Du sẵn sàng phá luật câu 6 chữ, thì Nguyễn Đình Chiểu và Nguyễn Bính hầu như tuân theo luật *bttb*.

Tuy nhiên, có lẽ không ngạc nhiên, cả ba nhà thơ đều cấu trúc câu 6 chữ theo luật chuẩn *bb tt bb*: khoảng 18% trong *Kiểu*, 14% trong *Lục Vân Tiên*, và 19% trong thơ Nguyễn Bính. Ngoài ra, Nguyễn Bính có khuynh hướng cấu trúc câu sáu chữ theo nhịp *bb bt bb* (19%), nhiều hơn *Kiểu* (chỉ 10%), và *Lục Vân Tiên* (13%). Một biến thể khác khá thông dụng trong thơ Nguyễn Bính là *tb tt bb* (18%), so với *Kiểu* (17%) và *Lục Vân Tiên* (14%).

Trong câu 8 chữ, sự khác biệt giữa ba nhà thơ còn đậm nét hơn. Chẳng hạn như trong tất cả các biến thể của câu tám chữ, Nguyễn Bính thường dùng cấu trúc chuẩn *bb tt bb tb* (11,3%) và *bb bt tb bb* (11,3%), so với 5% đến 7% trong *Kiểu* và *Lục Vân Tiên*. Ngược lại, Nguyễn Bính ít dùng nhưng biến thể mà hai nhà thơ lớn kia hay dùng như: *tb bt tb bb* (8% trong *Kiểu*, 9% trong *Lục Vân Tiên*, và 7% trong thơ Nguyễn Bính); *tb bt tb tb* (8,5% trong *Kiểu*, 6% trong *Lục Vân Tiên*, và 1,5% trong thơ Nguyễn Bính); hay *tb tt tb tb* (7,1% trong *Kiểu*, 6,4% trong *Lục Vân Tiên*, và 3% trong thơ Nguyễn Bính); hay *bb bt tb tb* (6% trong *Kiểu*, 6% trong *Lục Vân Tiên*, và 2% trong thơ Nguyễn Bính).

Một vài nhận xét sơ khởi

Phân tích một tác phẩm văn chương bằng toán thống kê là một việc làm tương đối mới, nếu không muốn nói là hơi ... ngô ng. Mặc dù thống kê học đã được ứng dụng thành công rực rỡ vào nhiều hoạt động hay nghiên cứu như kĩ thuật, kinh tế, y khoa, và thậm chí xã hội học, nhưng ứng dụng thống kê trong văn chương vẫn còn là một lãnh địa bỏ hoang. Trong lịch sử thống kê học, có hai công trình dùng

thống kê trong văn chương: một dùng trong việc phân tích từ ngữ trong tác phẩm *Ulysses* (của Nhà văn nổi tiếng James Joyce), và một công trình khác dùng thống kê để xác định tác giả của những bài viết *The Federalists* bên Mỹ. Có lẽ bài này là một ứng dụng thống kê học đầu tiên vào việc phân tích văn điệu ở nước ta.

Tuy nhiên, trong bài viết này, tôi chỉ trình bày một vài kết quả sơ lược của bản phân tích, đặc biệt chú trọng vào sự phân phối văn điệu. Phần hai của bài phân tích sẽ chú trọng vào các lĩnh vực toán phức tạp hơn và sẽ trình bày trong một bài viết khác. Vì đây là một phân tích theo dạng mô tả, cho nên không có suy luận gì đáng kể có thể rút ra từ kết quả. Tuy nhiên, người viết thử đưa ra vài nhận xét qua kết quả phân tích trên như sau:

Thứ nhất, một điều khá thú vị là, tính trung bình, những chữ trong tiếng Việt chỉ có 3,4 mẫu tự. Chữ dài nhất của ta cũng chỉ có 7 mẫu tự (chữ *Nghiêng*)! Trong khi đó, trong tiếng Anh, tính trung bình, mỗi chữ dài đến 4,6 mẫu tự, và có chữ dài hơn 10 mẫu tự. Điều thú vị thứ hai là độ lệch tiêu chuẩn (standard deviation) của từ tiếng Việt chỉ 0,93, rất thấp so với độ lệch trong tiếng Anh là 2,6. Điều này có ý nghĩa gì? Theo nhà thần kinh – tâm lý học (neuropsychologist) S. Dehaene thuộc Viện nghiên cứu y khoa Pháp (Paris) thì người nào dùng chữ ngắn thường có khả năng nhớ các con số lâu hơn. Chẳng hạn như người Tô Cách Lan (Scotland) có tên của những con số dài hơn, và thường có khả năng tính toán [tính trung bình] thấp hơn, người Anh. Nếu thuyết tương quan giữa độ dài của từ ngữ và năng khiếu tính toán là đúng, ta cũng có thể giải thích một phần tại sao trẻ em người Việt tương đối có năng khiếu về số học khá hơn trẻ em người Tây phương chẳng hạn. Tuy nhiên, có lẽ đó là một giả thuyết cần khai thác thêm.

Thứ hai, những khác biệt về mức độ phân phối các biến thể câu 6 và 8 chữ trong thơ giữa ba nhà thơ có thể cho ta một nhận xét chung là so với Nguyễn Du và Nguyễn Đình Chiểu, Nguyễn Bính là nhà thơ thích thanh bằng. Thực vậy, nếu tính theo số từ-duy-nhất, thơ lục bát của Nguyễn Bính dùng đến 55% từ thanh bằng, 11% cao cao hơn *Kiểu*

(44%) và 6% cao hơn *Lục Vân Tiên* (49%). Xác suất mà mức độ khác biệt này do các yếu tố ngẫu nhiên chỉ 0,001 phần trăm. Nói cách khác, xu hướng dùng từ thanh bằng của Nguyễn Bính có thể là một thói quen có ý thức, có hệ thống, chứ không phải do yếu tố ngẫu nhiên gây ra.

Thứ ba, phân tích sự khác biệt về nguyên âm và phụ âm giữa ba nhà thơ trên cũng cho thấy một vài khuynh hướng đáng chú ý. Chẳng hạn như về nguyên âm bổng (tức những *i, e, và ê*) không có sự khác biệt nào giữa ba nhà thơ (khoảng 20% đến 21%). Nhưng những chữ có nguyên âm trầm (*u, o, ô*) chiếm đến 35% từ ngữ trong thơ Nguyễn Bính, và đây cũng là một khác biệt có ý nghĩa thống kê (chứ không phải ngẫu nhiên) so với hai tác phẩm *Kiểu* và *Lục Vân Tiên* (khoảng 30 đến 31%).

Nguyễn Bính còn có khuynh hướng dùng phụ âm vang (như *m, n, nh, và ng*) cao hơn Nguyễn Du và Nguyễn Đình Chiểu. Trong thơ lục bát của Nguyễn Bính và Nguyễn Đình Chiểu, có đến 80% và 81% là những chữ có phụ âm vang, cao hơn so với *Kiểu* (77%). Ngược lại, *Nguyễn Du* có khuynh hướng dùng những chữ có phụ âm tắc (như *p, t, ch* và *c*), khoảng 23%, hơn Nguyễn Đình Chiểu (20%) và Nguyễn Bính (19%). Qua so sánh này, có thể nói Nguyễn Bính hình như thích dùng những chữ có phụ âm vang và nguyên âm trầm hơn Nguyễn Du và Nguyễn Đình Chiểu. Thực vậy, trong thơ Nguyễn Bính, có khoảng 28% là những chữ vang-trầm, so với 23% trong *Kiểu*, và 25% trong *Lục Vân Tiên*.

Thế thì một câu thơ lục bát mượt mà nhất nên được cấu trúc ra sao? Kết quả phân tích này không cho một câu trả lời dứt khoát, nhưng dựa vào những tần số sử dụng về vần điệu và âm điệu, một bài thơ lục bát “hay” có lẽ nên được cấu trúc theo luật *bb bt bb / bb tt bb tb, bb tt bb / bb bt tb bb, tb tt bb / bb bt bb bb*, và *tb bt bb / tb bt tb bb*. Ngoài ra âm điệu của chữ dùng có lẽ nên sử dụng những chữ có phụ âm vang và nguyên âm trầm càng nhiều càng tốt.

Tuy nhiên, tôi không tin là khi viết sáng tác, các nhà thơ đã ngồi xuống tính toán nên dùng bao nhiêu chữ thanh bằng, trắc, bao nhiêu chữ có nguyên âm vang, phụ âm tắc, v.v... mà có thể chỉ

vì một thói quen dùng chữ mà thôi. Nhưng cách dùng này chắc chắn không phải ngẫu nhiên.

Dĩ nhiên, việc phân tích một bản văn hay thơ, nhất là phân tích thống kê, không thể thay thế đọc bản văn đó. Kết quả phân tích trình bày trên đây cho thấy một số qui luật và từ ngữ được các nhà thơ sử dụng có thể góp phần vào việc nghiên cứu ngữ học. Một mục đích của bài phân tích này là phác họa một vài khả năng mà thống kê học có thể ứng dụng được trong việc phân tích từ ngữ trong thơ văn.

Có thể một vài kết quả sẽ làm ngạc nhiên nhiều bạn đọc. “Con số thống kê như như một bộ đồ tắm bikini: những gì nó biểu hiện ra ngoài thì hấp dẫn đấy; nhưng những gì nó che đậy mới là quan trọng.” Hi vọng bài viết này có thể khơi dậy sự tò mò của quý bạn đọc trẻ, và thúc đẩy việc đi tìm những ý nghĩa đằng sau đang bị che đậy bởi những con số thống kê./.

Đoàn Dũng

sưu tầm từ <http://www.statistics.vn>

Chú thích:

[1] Chẳng hạn như trong câu “Một, hai nghiêng nước nghiêng thành”, có tất cả là 6 từ đơn, nhưng chỉ 5 từ gốc (vì chữ *Nghiêng* được lặp lại 2 lần).

[2] Chẳng hạn như chữ “*Nghiêng*”, có 7 mẫu tự.

[3] Để hiểu ý nghĩa của chỉ số lệch tiêu chuẩn này, xin lấy một ví dụ sau đây như một lời giải thích. Giả sử ta có ba dãy số như sau: (a) 5, 5, 5, 5, 5; (b) 5, 4, 5, 5, 6; và (c) 2, 1, 1, 3, 18. Cả ba dãy số này đều có chỉ số trung bình là 5. Tuy nhiên, sự phân phối của các dãy số này khác nhau, vì quan sát kỹ qua những con số trong từng dãy số cho thấy mức độ biến thể hay “di động” của chúng rất khác nhau. Trong dãy số (a), hoàn toàn không có biến thể; trong dãy (b), mức độ biến thể tương đối thấp vì các con số phân phối chung quanh 4 và 6; và trong dãy (c), mức độ biến thể rất cao, đi từ 1 đến 18. Chỉ số lệch tiêu chuẩn dùng để đo mức độ biến thiên này. Công thức dùng để tính chỉ số lệch tiêu chuẩn là như sau: $\sqrt{[(x_1 - m)^2 + (x_2 - m)^2 + (x_3 - m)^2 + \dots + (x_n - m)^2] / (n-1)}$; trong đó “sqrt” có nghĩa là căn số bậc hai, $x_1, x_2, x_3, \dots, x_n$ là những con số riêng lẻ trong một dãy số, n là số lượng con số trong một dãy số, và m là số trung bình. Ví dụ như trong dãy số (c), ta tính như sau: $\sqrt{[(2-5)^2 + (1-5)^2 + (1-5)^2 + (3-5)^2 + (18-5)^2] / (5-1)} = 7.3$; trong dãy (b) $\sqrt{[(5-5)^2 + (4-5)^2 + (5-5)^2 + (5-5)^2 + (6-5)^2] / (5-1)} = 0.71$; và trong dãy (a), dĩ nhiên chỉ số lệch tiêu chuẩn là 0, vì các con số trong đó bất biến.

Bảng 1. Phân phối phụ âm và nguyên âm

Phụ âm	Nguyên âm	Kiểu (%)	Lục Vân Tiên (%)	Lỡ bước sang ngang (%)
Vang	Bổng	15,7	14,5	16,8
Vang	Trầm	23,2	24,7	27,6
Vang	Trung	40,0	42,5	38,2 **
Tắc	Bổng	1,8	1,6	1,2
Tắc	Trầm	9,0	7,1	9,0
Tắc	Trung	10,3	9,7	7,3 **
Tổng cộng		100	100	100

Những số được đánh dấu ** là những khác biệt có ý nghĩa thống kê (statistical significance) giữa ba tác phẩm. Khác biệt "có ý nghĩa thống kê" là mức độ khác nhau không phải do các yếu tố ngẫu nhiên gây nên, mà có thể do thói quen dùng từ.

Bảng 2. Phân phối vần trong câu 6 chữ

Vần 6 câu	Kiểu	Lục Vân Tiên	Lỡ bước sang ngang
<i>b b b b t b</i>	0	0,1	0,8
<i>b b b t b b</i>	10,4	12,9	18,8 **
<i>b b b t t b</i>	9,7	14,9	10,5 **
<i>b b t b b b</i>	0,1	0,1	0
<i>b b t b t b</i>	0,1	0,1	0
<i>b b t t b b</i>	17,6	13,7	18,8
<i>b b t t t b</i>	8,9	11,1	3,8 **
<i>b t b t b b</i>	0,1	0	0
<i>b t b t t b</i>	0,1	0	0
<i>b t t b b b</i>	0,1	0	0
<i>b t t b t b</i>	0,2	0	0
<i>b t t t b b</i>	0,6	0	0
<i>b t t t t b</i>	0,1	0	0
<i>t b b b b b</i>	0,1	0	0
<i>t b b b t b</i>	0	0	0
<i>t b b t b b</i>	12,1	12,3	13,5
<i>t b b t t b</i>	13,3	12,1	11,3
<i>t b t b b b</i>	0,1	0	0,8

Vần 6 câu	Kiểu	Lục Vân Tiên	Lỡ bước sang ngang
<i>t b t t b b</i>	16,6	14,3	18,0 **
<i>t b t t t b</i>	9,6	8,3	1,5 **
<i>t t b b b b</i>	0	0	1,5
<i>t t b t b b</i>	0,1	0	0
<i>t t b t t b</i>	0,1	0	0
<i>t t t t b b</i>	0,1	0	0
<i>t t t t t b</i>	0,1	0,1	0

Những số được đánh dấu ** là những khác biệt có ý nghĩa thống kê giữa ba tác phẩm.

Bảng 3. Phân phối vần trong câu 8 chữ

Vần 8 câu	Kiểu	Lục Vân Tiên	Lỡ bước sang ngang
<i>b b b t b b b b</i>	5,1	4,9	9,1 **
<i>b b b t b b t b</i>	5,7	5,0	6,0
<i>b b b t t b b b</i>	5,7	6,8	11,3 **
<i>b b b t t b t b</i>	6,4	5,7	2,3 **
<i>b b t t b b b b</i>	4,4	5,7	6,8
<i>b b t t b b t b</i>	7,2	6,2	11,3 **
<i>b b t t t b b b</i>	5,7	6,8	9,0 **
<i>b b t t t b t b</i>	6,3	6,4	6,0
<i>b t t t t b t b</i>	0	0,1	1,5
<i>t b b t b b b b</i>	4,3	6,1	3,8
<i>t b b t b b t b</i>	7,0	5,0	5,3
<i>t b b t t b b b</i>	8,0	8,8	6,8
<i>t b b t t b t b</i>	8,5	6,0	1,5 **
<i>t b t t b b b b</i>	3,5	5,8	4,5
<i>t b t t b b t b</i>	8,4	6,6	6,8
<i>t b t t t b b b</i>	6,9	7,6	5,3
<i>t b t t t b t b</i>	7,1	6,4	3,1 **
<i>t t b b b t b b</i>	0	0,1	0
<i>t t b t b b b b</i>	0	0,1	0
<i>t t t t b b t b</i>	0	0,1	0