

VÀI KẾT QUẢ KHẢO SÁT SƠ BỘ VỀ TỪ VỰNG TIẾNG VIỆT CỔ TRONG MỘT SỐ TÁC PHẨM NÔM

PGS.TS VŨ ĐỨC NGHIÊU

Tiếng Việt cổ ở đây được hiểu là tiếng Việt trong khoảng thời gian từ thế kỉ XIII đến thế kỉ XVI theo sự phân định và đề xuất của Nguyễn Tài Căn trong một bảng phân kì lịch sử tiếng Việt [12] căn cứ trên những nhân tố lịch sử xã hội, ngôn ngữ, văn hoá và văn tự... Nghiên cứu những chứng tích từ vựng thuộc giai đoạn này chắc chắn sẽ có thêm được cứ liệu để góp phần vào việc khảo tả lịch sử từ vựng nói riêng và lịch sử tiếng Việt nói chung.

Đặt vấn đề như vậy, chúng tôi đã khảo sát trong 10 nguồn ngữ liệu thành văn cụ thể sau đây:

1) *Phật thuyết đại báo phụ mẫu ân trọng kinh* [5] (gọi tắt là *Phật thuyết*)¹.

2) *Cư trần lạc đạo phú*, *Đặc thú lâm tuyền thành đạo ca*, *Vịnh Hoa Yên tự phú*, *Giáo từ phú* [3].

3) *Thiền tông Khoá hư ngữ lục* (gọi tắt: *Khoá hư lục*) khoảng cuối thế kỉ XIV [7].

4) *Quốc âm thi tập* - tập thơ nôm của Nguyễn Trãi, thế kỉ XV [6].

5) *Hồng Đức quốc âm thi tập* thế kỉ XV [4].

6) *Thơ Nôm của Nguyễn Bỉnh Khiêm* thế kỉ XVI [8].

7) *Chi Nam ngọc âm giải nghĩa* (trước thế kỉ XVII, không sớm hơn thế kỉ XIV [2]).

8) *Việt sử diễn âm* (giữa thế kỉ XVI) [10].

9) *Truyền kì mạn lục giải âm* (thế kỉ XVI) [9].

10) *Hoa Di dịch ngữ* (khoảng thế kỉ XV - XVI) [1]

Dưới đây là một số kết quả cụ thể.

1. Về cấu trúc hình thái của từ

1.1. Các nguồn ngữ liệu được khảo sát đều phản ánh rất rõ ràng: Vào giai đoạn tiếng Việt cổ, (và chắc trong giai đoạn tiếng tiền Việt trở về trước lại càng như thế) có rất nhiều từ còn giữ hình thái cổ. Cụ thể là có ba biểu hiện sau đây:

1.1.1. Thứ nhất, nhiều từ còn giữ cấu trúc song tiết (sesquisyllable) từ thời Proto Việt Mường, gồm một tiền âm tiết (còn gọi là âm tiết phụ) và một âm tiết chính [CvCVC]. Vào thời kì ra đời của các nguồn ngữ liệu đang xét, chứng tích của các cấu trúc song tiết đó là ở cách ghi một từ Nôm (Việt) bằng hai mã chữ Hán. Thí dụ, trong *Phật thuyết*..., chúng tôi thấy có các dấu vết tiền âm tiết (xem [13], [14], [15], [16]):

*[a] được ghi bằng chữ Hán Việt "a". Thí dụ: *bãm* (a - biếm), *như* (a - dư), *đơ* (a - đơ), *giết gia* (a - chất a - chí), *gây* (a - kέ), *lừa* (a - lā), *này* (a

- ni), *sánh* (a - *tingh*), *vội* (a - *bôi*), *vỡ* (a - *phó*), *vỗ về* (a - *phố a - phê*)...

*[pə] được ghi bằng chữ Hán Việt “bà”. Thí dụ: *đắp* (ba - *đạp*), *đề* (ba - *đê*), *đứt* (ba - *đất*), *bói* (ba - *lai*), *nái* (ba - *nãi*), *dượt* (ba - *dục*), *sung* (ba - *lăng*), *ran* (ba - *tan*)...

*[p'ə] được ghi bằng chữ Hán Việt “pha/ phá”. Thí dụ: *trật* (pha - *lật*), *rắn* (phá - *tán*), *sáu* (phá - *liễu*), *sốt* (phá (pha) - *luật*)...

*[cə] được ghi bằng chữ Hán Việt “xa / xá”. Thí dụ: *màng/ mắng* (xa/ xá - *mãng*), *cho* (xá - *chu*), *quàng* (xá - *quang*), *mở* (xá - *mĩ*), *thốt* (xá - *thuyết*), *xét* (xá - *sát*)...

*[k'ə] được ghi bằng chữ Hán Việt “khả”. Thí dụ: *nhớ* (khả - *nhữ*), *sắt* (khả - *liệt*), *xa* (khả - *gia/ xa*).

*[tə] được ghi bằng chữ Hán Việt “đa”. Thí dụ: *bán* (đa - *bản*), *cưỡi* (đa - *cái*), *bên* (đa - *biên*), *ghê* (đa - *kì*), *mình* (đa - *mệnh*), *lung* (đa - *lăng*), *mệt* (đa - *miệt*), *nghe* (đa - *nghi*), *vai* (đa - *mai*), *mắc mớ* (đa - *mặc đa - phủ*)...

*[lə] được ghi bằng chữ Hán Việt “la”. Thí dụ: *đá* (la - *đả*), *ngàn* (la - *ngạn*)...

*[sə] được ghi bằng chữ Hán Việt “tư”. Thí dụ: *vua* (tư - *bố*), *vui* (tư - *bui*)...

Các từ song tiết (song âm) dạng [CvCVC] như vừa trình bày trên đây là một trong những biểu hiện về mặt hình thái học cần phải được chú ý thích đáng trong khi nghiên cứu tiếng Việt cổ. Khi nhìn theo thời gian trở xuôi, nếu trong văn bản *Phật thuyết*, những cấu trúc từ có tiền âm tiết xuất hiện

khá nhiều, với tần số cao (theo khảo sát của Nguyễn Quang Hồng [14] có 108 lần xuất hiện cấu trúc CvCVC), thì trong các văn bản sau đó (*Khoá hư lục*, *Quốc âm thi tập*, *Chi Nam ngọc âm giải nghĩa*) cả số lượng lẫn tần số xuất hiện của chúng đều giảm hẳn. Trong *Khoá hư lục*, ta có gặp từ *bà cắt*; trong *Quốc âm thi tập*, ta có gặp từ: *la đá*, *la ngàn*, *bà ngựa*; còn trong *Chi Nam ngọc âm giải nghĩa* - văn bản được coi là một quyển từ điển cổ khá lớn, gồm 40 chương, 3.394 mục từ thì mới gặp nhiều hơn những hình thức ghi bằng chữ nôm có thể ngờ là phản ánh cấu trúc CvCVC như: (*bà*) *bồ cắt*, *bồ câu*, *bồ cò*, *bồ cu*, *bồ cóc*, *bồ hòn*, *bồ nâu*, *bồ đài*, *bồ kết*, *bồ nâu*, *bồ nông*, *bồ ngư*, *bồ cóc*, *bồ ván*, *bệnh rết*, *la đá*, *lỗ vừng*, *lỗ mướp*, *lú bú*, *lù và*...

1.1.2. Thứ hai, “mặc dù chịu ảnh hưởng sâu đậm từ cơ chế đơn tiết của tiếng Hán và chữ Hán, người Việt thuở ấy đôi khi cũng đã Việt hoá một số từ đơn tiết Hán theo cách phát âm CvCVC hoặc CCVC trong bản ngữ của mình” [14, 141-142]. Chữ Nôm trong văn bản *Phật thuyết* phản ánh khá rõ hiện tượng này. Thí dụ:

Dùng chữ Hán Việt *tingh*, nhưng viết hai chữ [a - *tingh*] để ghi từ *sánh*

Dùng chữ Hán Việt *bản*, nhưng viết hai chữ [đa - *bản*] để ghi từ *bán*

Dùng chữ Hán Việt *biên*, nhưng viết hai chữ [đa - *biên*] để ghi từ *bên/ biên*.

Dùng chữ Hán Việt *mệnh*, nhưng viết hai chữ [đa - *mệnh*] để ghi từ (*thân*) *mình*

Dùng chữ Hán Việt *hằng*, nhưng viết hai chữ [cá - hằng] để ghi từ *hằng* (=thường)

Dùng chữ Hán Việt *bằng*, nhưng viết hai chữ [đa - bằng] để ghi từ *bằng*

Dùng chữ Hán Việt *quyển*, nhưng viết hai chữ [đa - quyển] để ghi từ *quyển*

1.1.3. Thứ ba, rất nhiều từ còn giữ tổ hợp phụ âm, phản ánh cấu trúc âm tiết [CCVC] vốn có từ những giai đoạn cổ xưa hơn. Tư liệu trong các văn bản viết bằng chữ Nôm, trong *Hoa Di dịch ngữ* đều phản ánh rất rõ điều này: ghi một từ bằng hai mã chữ, mỗi mã chữ thể hiện một C (phụ âm) trong tổ hợp âm đầu CC².

Thí dụ, trong *Phật thuyết* [13, 15-16] còn lưu giữ:

- Tổ hợp *[bl]: *trá* ghi bằng (ba-la), *trái* (ba - lai), *trợn* (bà - luận), *trớ* (ba - lữ), *trời* (ba - lệ), *trối* (bà - lỗi), *tràn* (ba - lam), *trát* (ba - lạt)...

- Tổ hợp *[p'1]: *sốt* ghi bằng (pha-luật), *sáu* (phá - liễu), *trật* (pha - lật)...

- Tổ hợp *[k'1]: *sắt* (khả - liệt), *sa* (khả - la)...

- Tổ hợp *[kl]: *trông* (cự/ cá - lung), *trăm* (cá - lâm), *trưa* (cá - lư), *lòng* (cá - lộng), *trẻ* (cá - lể)...

- Tổ hợp *[ml]: *lăn lóc* (ma - lân ma - lộc), *lận* (ma - lận), *lớn* (ma - lận), *lời* (ma - lệ), *giận* (ma - lận)...

- Tổ hợp *[kr]: *sống* (cổ - lộng), *sữa* (cự - lữ), *sung* (cự - lảng)...

Trong các nguồn ngữ liệu khác như: *Cư trần lạc đạo phú*, *Khoá hư lục*, *Quốc âm thi tập*, *Hồng Đức quốc âm thi tập*, *Thơ Nôm của Nguyễn Bình Khiêm*, *Chi Nam ngọc âm giải*

ngĩa, *Việt sử diễn âm*, *Truyền kì mạn lục giải âm*... cũng có tình hình tương tự: các tổ hợp phụ âm vẫn còn tồn tại và được thể hiện qua chứng tích chữ Nôm. Thí dụ:

*[bl-] ở chữ (*ba - lai*) = *trái*, (*bà*) = *trả*... *[kl-] ở chữ (*công*) = *trong*, (*cư - lược*) = *trước*, (*cư - lâu*) = *sau*, (*cự - lang*) = *sang*... *[ps-] ở chữ (*tu - bôi*) = *vui*... (*Cư trần lạc đạo phú*).

*[kl-] ở chữ (*cổ - lộng*) = *trống*, (*cự - lược*) = *trước*, (*công / cộng*) = *trong*, (*lược*) = *trước*... *[kr-] ở chữ (*cự - lang*) = *sang*, (*cổ - la*) = *sa*, (*cư - lô*) = *so*... (*Khoá hư lục*).

*[bl-] ở chữ (*ba-lai*) = *trái*, (*ba-lăng*) = *trăng*, (*bà*) = *trả*... *[ml-] ở chữ (*miễn*) = *lãn*, (*ma-lê*) = *lời*... *[kl-] ở chữ (*cự-liêu*) = *treo*, (*cự-lược*) = *trước*, (*cổ-lộng*) = *trống*, (*cư-lung*) = *trông*, (*cự-luân*) = *tròn*... *[kr-] ở chữ (*cự-lôi*) = *soi*, (*cự-lang*) = *sang*, (*cổ-lộng*) = *sống*... *[ps-] ở chữ (*tu-bôi*) = *bui*, (*tu-bôi*) = *vui*... (*Quốc âm thi tập*).

*[kl-] ở cách ghi âm *trâu* = (*cách - lâu*), *trống* = (*cộng*), gặp *[tl-] ở cách ghi âm *trán* = (*thác-lan*), *trong* = (*lộng*)... *[pl-], *[bl-] ở cách ghi âm *trái* = (*bái*), *trai* = (*lai*), *trầu* = (*lâu*), *trời* = (*lối*)... (*Hoa Di dịch ngữ*).

*[bl-] ở chữ (*bà*) = *trả*, (*ba-lăng*) = *trăng*, (*ba-lan*) = *trán*... *[ml-] ở chữ (*mãnh-lâu*) = (*cá*) *lầu*... *[kl-] ở chữ (*cổ-lộng*) = *trong*, (*cự-luân*) = *tròn*, (*cư-lô*) = *trò* (*chuyện*)... *[kr-] ở chữ (*cự-lập*) = (*cái*) *sập*, (*cư-lâm*) = *sám*... *[tl-] ở chữ *tu-lai* = (*ốc*) *trai*... (*Chi Nam ngọc âm giải nghĩa*).

Cứ liệu về các tổ hợp phụ âm như vừa trình bày trên đây, cho đến nay, đã được giới nghiên cứu hữu quan nhắc đến khá nhiều. Việc dùng hai mã chữ Hán để “phiên âm”/ ghi âm một từ Việt thời Việt cổ, không có gì xa lạ; người Việt đã làm, và người phương Bắc cũng đã từng làm³.

Cho đến *Truyền kì mạn lục giải âm* (thế kỉ XVI) các sự kiện tương tự vẫn tiếp tục thấy được như vậy⁴.

1.2. Về tương quan giữa các đơn vị từ vựng đơn tiết với đơn vị đa tiết.

Bảng 1. Số đơn vị từ vựng đơn tiết, đa tiết trong các nguồn ngữ liệu được khảo sát.

Nét đối lập đơn tiết - đa tiết của từ nói riêng, của đơn vị từ vựng nói chung, luôn luôn là một đặc điểm nổi bật và được quan tâm miêu tả trong các nghiên cứu về tiếng Việt. Nếu căn cứ vào kết quả phiên âm theo cách nhìn đồng đại ngày nay (tạm gác các đơn vị đa tiết “nguyên thủy” có cấu trúc [CvCVC] như vừa trình bày trên đây sang một bên; vì nếu có tính đến thì số này cũng không làm thay đổi đánh giá chung tương quan về lượng của chúng), chúng tôi thu được số liệu sau đây⁵:

NGUỒN NGỮ LIỆU (VĂN BẢN)	SỐ ĐƠN VỊ ĐƠN TIẾT	SỐ ĐƠN VỊ ĐA TIẾT	TỔNG
<i>Phật thuyết</i> ⁶	922 (83,0%)	187 (17,0%)	1.109(100%)
<i>Bốn bài phú đời Trần</i>	996 (76,6%)	304 (23,4%)	1.300(100%)
<i>Quốc âm thi tập</i> ⁷	1.553(73,6%)	557 (26,4%)	2.110(100%)
<i>Hồng Đức quốc âm thi tập</i> ⁸	1.953 (63,8%)	1.108(36,2%)	3.061(100%)
<i>Hoa Di dịch ngữ</i> ⁹	510 (92,7%)	40 (7,3%)	550 (100%)
<i>Thơ nôm Ng. Bình Khiêm</i> ¹⁰	1.278 (76,0%)	403 (24,0%)	1.681(100%)

Bảng 1 cho thấy trong các nguồn ngữ liệu được khảo sát, tỉ lệ từ đơn tiết là tỉ lệ chiếm ưu thế. Trong sáu nguồn được kiểm đếm, tỉ lệ từ đơn thấp nhất như ở *Hồng Đức quốc âm thi tập* (HĐQATT) cũng là trên 63% (thực ra, nếu phân giới các đơn vị từ vựng trong tác phẩm này một cách “sát” hơn nữa, thì tỉ lệ đó chắc chắn sẽ cao hơn); còn trong *Quốc âm thi tập* (QAATT), *Thơ nôm Nguyễn Bình Khiêm*, *Bốn bài phú đời Trần*, *Phật thuyết* đều trên 70%¹¹; thậm chí, trong *Hoa Di dịch ngữ* là trên 90%.

Một số liệu khác phản ánh mức độ được sử dụng (hoặc cũng có thể coi như năng lực hoạt động) của các đơn vị từ vựng đơn tiết và đa tiết trong các nguồn ngữ liệu đang xét được thể hiện qua bảng sau đây:

Bảng 2. Tương quan tổng tần số của các từ (đơn vị từ vựng) đơn tiết, đa tiết với tổng độ dài văn bản được khảo sát.

NGUỒN NGỮ LIỆU (VĂN BẢN)	TỔNG TẦN SỐ CÁC ĐƠN VỊ ĐƠN TIẾT	TỔNG TẦN SỐ CÁC ĐƠN VỊ ĐA TIẾT	ĐỘ DÀI VĂN BẢN (TỔNG TẦN SỐ CÁC TỪ)
<i>Phật thuyết</i> ¹²	4.772 (91,9%)	428 (8,2%)	5.200(100%)
Bốn bài phú đời Trần	2.535 (87,0%)	379 (13,0%)	2.914(100%)
<i>Quốc âm thi tập</i> ¹³	10.164 (90,7%)	1.044 (9,3%)	11.208(100%)
<i>Hồng Đức quốc âm thi tập</i>	13.506 (88,8%)	1.701 (11,2%)	15.207(100%)
<i>Thơ nôm Ng. Bình Khiêm</i>	5.674 (90,4%)	599 (9,6%)	6.273 (100%)

Bảng 2 cho thấy rõ tình thế áp đảo của các đơn vị từ vựng đơn tiết so với các đơn vị đa tiết cả về mặt số lượng trong vốn từ lẫn mặt chiếm lĩnh độ dài văn bản (số lần được sử dụng của chúng trong các văn bản được khảo sát).

So sánh bảng 1 và bảng 2 với nhau, ta sẽ thấy: nếu như trong các nguồn ngữ liệu đó, số lượng các đơn vị từ vựng đơn tiết cao hơn số đơn vị đa tiết khoảng hai đến ba lần: (Bốn bài phú đời Trần: 996/ 304; *Quốc âm thi tập*: 1554/ 557; *Hồng Đức quốc âm thi tập*: 1953/1108; *Thơ Nôm Nguyễn Bình Khiêm*: 1278/ 403); cùng lắm là tới khoảng gần năm lần như trong *Phật thuyết*: 922/188; thì tổng tần số sử dụng của các đơn vị đơn tiết đó lại cao hơn tổng tần số sử dụng của các đơn vị đa tiết tới năm lần hoặc tám, thậm chí tới gần mười lần; cụ thể là:

<i>Phật thuyết</i> :	4.772 / 428	(11 lần)
Bốn bài phú đời Trần:	2.535 / 379	(6,7 lần)
<i>Quốc âm thi tập</i> :	10.164 / 1.044	(9,7 lần)
<i>Hồng Đức quốc âm thi tập</i> :	13.506 / 1.701	(7,9 lần)
<i>Thơ nôm Ng. Bình Khiêm</i> :	5.674 / 599	(9,5 lần)

Điều này cung cấp thêm một thông tin và góp thêm một căn cứ để nói rằng: Trong giai đoạn tiếng Việt cổ, các nguồn tư liệu được khảo sát cho thấy: các từ (đơn vị từ vựng) đơn tiết chiếm ưu thế hơn hẳn, thậm chí, có thể nói là áp đảo cả về số lượng trong vốn từ lẫn tần số sử dụng (chiếm lĩnh độ dài văn bản) khi đi vào hoạt động.

2. Về thành phần từ vựng phi Hán Việt và Hán Việt

Về mặt nguồn gốc, từ vựng tiếng Việt là kết quả của sự tích hợp các từ ngữ từ nguồn gốc bản địa, căn bản là Môn Khme, Tày Thái... và từ nguồn gốc Hán do những mối quan hệ láng giềng, tiếp xúc, vay mượn và ảnh hưởng, vừa vô cùng sâu sắc, vừa hết sức phức tạp. (Bộ phận từ ngữ có nguồn gốc

Ân Âu được vay mượn trong giai đoạn rất muộn về sau này, tiếng Việt cổ chưa hề biết đến).

Mặc dù là một kho tàng từ vựng có nhiều bộ phận xuất xứ từ những nguồn gốc khác nhau, nhưng đến giai đoạn tiếng Việt cổ, nét đối lập và đường ranh giới nổi trội trong nội bộ của nó không phải là đối lập giữa các nguồn gốc Môn Khme - Tây Thái - Nam Đảo - Hán, mà là phi Hán Việt với Hán Việt, trong đó phần phi Hán Việt bao gồm các bộ phận từ ngữ bản địa cùng với những từ ngữ gốc Hán đã nhập hệ sâu vào tiếng Việt (cổ Hán Việt, Hán Việt Việt hoá¹⁴ - bộ phận từ vựng mà mọi ứng xử ngôn ngữ của chúng đã hoàn toàn hoặc gần hoàn toàn như các từ bản địa). Đặt các từ ngữ Hán Việt trong tương quan với các từ ngữ phi Hán Việt để quan sát và phân tích, chúng tôi, trước hết khảo sát về mặt định lượng và có được một vài số liệu sau đây:

Bảng 3. Tương quan về lượng giữa từ Hán Việt, từ phi Hán Việt với tổng số từ của nguồn ngữ liệu được khảo sát

NGUỒN NGỮ LIỆU (VĂN BẢN)	SỐ TỪ NGỮ HÁN VIỆT	SỐ TỪ NGỮ PHI HÁN VIỆT	TỔNG
<i>Phật thuyết</i>	326 (29,4%)	783 (70,6%)	1109 (100%)
<i>Bốn bài phú đời Trần</i>	534 (41,0%)	766 (59,0%)	1300 (100%)
<i>Quốc âm thi tập</i>	823 (39,0%)	1287 (61,0%)	2110 (100%)
<i>Hồng Đức quốc âm thi tập</i>	1194 (39,0%)	1867 (61,0%)	3061 (100%)
<i>Hoa Di dịch ngữ</i>	181 (33,0%)	369 (67,0%)	550 (100%)
<i>Thơ nôm Ng. Bình Khiêm</i>	536 (32,0%)	850 (61,3%)	1681 (100%)

Bảng 3 cho thấy hai điểm đáng chú ý:

Thứ nhất, tỉ lệ từ ngữ phi Hán Việt trong các mẫu nghiên cứu lớn hơn nhiều so với từ ngữ Hán Việt. Cụ thể là chúng chiếm khoảng trên/ dưới hai phần ba vốn từ (gần 60% đến 70%), còn các từ Hán Việt chỉ chiếm trên/ dưới một phần ba (quãng xấp xỉ 30% đến 41%).

Thứ hai, tỉ lệ từ ngữ Hán Việt trên đây tương đối đồng đều giữa các mẫu nghiên cứu; và dựa theo đó, có thể quy các mẫu về ba nhóm:

- Nhóm thứ nhất: *Quốc âm thi tập* và *Hồng Đức quốc âm thi tập* cùng có tỉ lệ 39%;

- Nhóm thứ hai gồm: *Hoa Di dịch ngữ*, *Thơ Nôm Nguyễn Bình Khiêm* và *Phật thuyết* có tỉ lệ từ hơn 29% đến 33%;

- Nhóm thứ ba, riêng bốn bài phú đời Trần có tỉ lệ tới 41%.

Sự chênh lệch về các tỉ lệ như thế, có thể hiểu được. Nếu cân nhắc đến độ dài và đặc điểm thể loại của các văn bản (diễn ngôn) thì ta sẽ thấy độ chênh

về tỉ lệ giữa ba nhóm sẽ không còn là điều đáng phải băn khoăn nữa. Văn bản *Phật thuyết* có tỉ lệ từ Hán Việt thấp nhất (29,4%) có lẽ vì đây là bản dịch từ Hán văn nên các dịch giả đã rất cố gắng tìm kiếm những từ ngữ quen thuộc hơn, dễ hiểu hơn, (là các từ ngữ phi Hán Việt), để diễn giải các nội dung của bản kinh, nhằm làm cho việc phổ biến, truyền kinh được thuận lợi, dễ dàng hơn. Riêng bốn bài phú đời Trần thì tình hình lại khác. Tỉ lệ từ Hán Việt trong bốn bài phú này cao, vì độ dài văn bản không lớn (tổng độ dài bốn bài chỉ xấp xỉ tổng độ dài của văn bản *Phật thuyết*). Mặt khác, quan trọng hơn, chính là ở chỗ: trừ bài *Giáo từ phú* (cũng nói về cách sống, tu thân theo Phật giáo, về nhân quả tội lỗi và hình phạt theo cách hình dung của Phật giáo) không rõ được tác giả; còn lại, bài *Cư trần lạc đạo* và *Đặc thú lâm tuyền thành đạo ca* là của vua Trần Nhân Tông, bài *Hoa Yên tự phú* được tin là của Trúc lâm đệ tam tổ Huyền Quang; đây vốn là hai bậc đại nho, uyên bác về Hán học, tự viết để diễn giải về tâm đạo tu hành Phật giáo, để ca ngợi, miêu tả các cảnh chùa chiền, Phật sự... nên họ phải dùng nhiều từ ngữ Hán Việt để diễn đạt các khái niệm, các đặc ngữ... của Phật giáo, cũng là điều bình thường.

Về năng lực hoạt động (được sử dụng) trên văn bản, tương quan giữa các từ ngữ Hán Việt với phi Hán Việt được thể hiện trong bảng dưới đây:

Bảng 4. Tương quan giữa tần số sử dụng từ Hán Việt, từ phi Hán Việt với tổng độ dài văn bản của nguồn ngữ liệu được khảo sát

NGUỒN NGỮ LIỆU (VĂN BẢN)	TỔNG TẦN SỐ TỪ NGỮ PHI HÁN VIỆT	TỔNG TẦN SỐ TỪ NGỮ HÁN VIỆT	TỔNG ĐỘ DÀI VĂN BẢN
<i>Phật thuyết</i>	4.229 (81,3%)	971 (18,7%)	5.200 (100%)
<i>Bốn bài phú đời Trần</i>	1.925 (66,0%)	989 (34,0%)	2.914 (100%)
<i>Quốc âm thi tập</i>	8.021 (71,6%)	3.187 (28,4%)	11.208 (100%)
<i>Hồng Đức quốc âm thi tập</i>	11.031 (72,5%)	4.176 (27,5%)	15.207 (100%)
<i>Thơ Nôm Ng. Bình Khiêm</i>	4.887 (77,0%)	1.386 (23,0%)	6.273 (100%)

Bảng 4 cho thấy một tình hình rất rõ: không chỉ chiếm số đông áp đảo trong vốn từ của các nguồn ngữ liệu được khảo sát, các từ phi Hán Việt còn thể hiện năng lực hoạt động cao của mình trong việc chiếm lĩnh độ dài của các văn bản. So với phần độ dài văn bản do các từ ngữ Hán Việt chiếm lĩnh (tổng tần số của chúng), rõ ràng, phần văn bản do các từ ngữ phi Hán Việt chiếm lĩnh là phần có tỉ lệ áp đảo. So sánh:

Trong văn bản <i>Phật thuyết</i> :	81,3% - 18,7%
Trong văn bản bốn bài phú đời Trần:	66,0% - 34,0%
Trong văn bản <i>Quốc âm thi tập</i> :	71,6% - 28,4%
Trong văn bản <i>Hồng Đức quốc âm thi tập</i> :	72,5% - 27,5%
Trong văn bản <i>Thơ Nôm Ng. Bình Khiêm</i> :	77,0% - 23,0%.

Ngoài những thông tin trên đây, năng lực hoạt động của các từ ngữ Hán Việt và phi Hán Việt còn được thể hiện qua tần số sử dụng (xuất hiện) trung bình của chúng trong các nguồn tư liệu. Điều này có thể kiểm chứng được qua các số liệu sau:

Bảng 5. Tần số trung bình của các từ ngữ Hán Việt

NGUỒN NGỮ LIỆU	SỐ TỪ NGỮ HÁN VIỆT	TỔNG TẦN SỐ TỪ NGỮ HÁN VIỆT	TẦN SỐ TRUNG BÌNH
<i>Phật thuyết</i>	326	971	2,97
<i>Bốn bài phú đời Trần</i>	534	989	1,85
<i>Quốc âm thi tập</i>	823	3.178	3,87
<i>Hồng Đức quốc âm thi tập</i>	1194	4.167	3,48
<i>Thơ Nôm Ng. Bình Khiêm</i>	536	1386	2,58

Bảng 6. Tần số trung bình của các từ ngữ phi Hán Việt

NGUỒN NGỮ LIỆU (VĂN BẢN)	SỐ TỪ NGỮ PHI HÁN VIỆT	TỔNG TẦN SỐ TỪ NGỮ PHI HÁN VIỆT	TẦN SỐ TRUNG BÌNH
<i>Phật thuyết</i>	783	4229	5,40
<i>Bốn bài phú đời Trần</i>	766	1925	2,51
<i>Quốc âm thi tập</i>	1287	8021	6,23
<i>Hồng Đức quốc âm thi tập</i>	1867	11040	5,91
<i>Thơ Nôm Ng. Bình Khiêm</i>	1145	4887	4,26

So sánh trong từng nguồn tư liệu được khảo sát ta thấy: ở *Phật thuyết*, tần số trung bình của từ ngữ phi Hán Việt lớn hơn tần số trung bình của từ ngữ Hán Việt là 1,8 lần (5,40 so với 2,97); ở *Bốn bài phú đời Trần*: 1,4 lần; ở *Quốc âm thi tập*: 1,6 lần; ở *Hồng Đức quốc âm thi tập*: 1,7 lần; ở *Thơ Nôm Nguyễn Bình Khiêm*: 1,7 lần.

Các số liệu phân tích định lượng này cho phép chúng ta khẳng định chắc được rằng: tuy các từ Hán Việt đóng vai trò vô cùng quan trọng trong kho tàng từ vựng tiếng Việt, nhưng trong các nguồn ngữ liệu được khảo sát, ưu thế về số lượng và năng lực hoạt động ngôn ngữ (thể hiện ở độ dài văn bản và tần số trung bình) lại không thuộc về chúng, mà thuộc về các từ ngữ phi Hán Việt. Điều này rất đáng được lưu ý vì trong bối cảnh ngôn ngữ, xã hội

đương thời, ảnh hưởng của tiếng Hán và từ vựng Hán Việt vô cùng sâu sắc và mạnh mẽ, từ vựng văn hoá, văn học tiếng Việt mới đang bắt đầu hình thành và phát triển, nhưng chúng ta vẫn thấy một quang cảnh từ vựng thể hiện trên ngữ liệu như vậy.

3. Về các từ ngữ cổ

3.1. Theo cách hiểu thường gặp, từ cổ là những từ đã từng tồn tại trước đây nhưng hiện nay không còn trong đời sống ngôn ngữ toàn dân đương đại nữa, hoặc nếu còn thì cũng đã có những biến đổi ngữ âm và/ hoặc ngữ nghĩa nhất định, hoặc chỉ còn làm thành tố trong những kết cấu ngôn ngữ rất hạn chế nào đó.

Trên thực tế, các biểu hiện và sự tồn tại của từ cổ khá đa dạng.

a. Thứ nhất, chúng có thể là những từ xưa đã từng tồn tại, hoạt động bình thường, nhưng nay đã hoàn toàn vắng mặt trong ngôn ngữ đương đại. Thí dụ: *mưa, bụi, thừa, bợ, phô, hoà, nhĩn, chung...*

b. Thứ hai, chúng có thể là những từ mà so với trước đây, đã có những biến đổi về mặt ngữ âm vì nhiều lí do trong quá trình phát triển lịch sử. Về căn bản, nghĩa của dạng cổ và dạng hiện nay của những từ như thế, không khác nhau. Thí dụ: *la đá - đá, lỗ vùng - vùng, lỗ mướp - mướp, bỏ cóc - cóc, bà (bò) cắt - cắt...*

c. Thứ ba, chúng có thể là những từ nay chỉ còn được dùng rất hạn chế trong phương ngữ nào đó với tư cách là từ ngữ địa phương, hoặc tồn tại trong một vài thành ngữ, tục ngữ nhất định. Thí dụ: bữa *rây/ rày, min*, ăn không ngồi *rồi*,...

d. Thứ tư, có thể là những từ xưa đã từng tồn tại, hoạt động bình thường, nhưng nay chỉ còn là những yếu tố đã “mất nghĩa” trong các đơn vị từ vựng như: hỏi *han*, hơn *nữa*, rõ *rệt*, lo *âu*, e *lệ*, *ngặt* nghèo...

e. Thứ năm, có thể là những từ trước đây được dùng với nghĩa cổ (và những đặc điểm ngữ pháp khác); nhưng nay các nghĩa cổ và/ hoặc những đặc điểm ngữ pháp đó không còn trong tiếng Việt hiện đại nữa. Thí dụ: từ *cái* dùng với nghĩa cổ như trong các kết cấu cú pháp: *cái rần, cái rùa, cái đê he, cái ve ve, cái quít chè...* từ *con* dùng với nghĩa cổ như trong các kết cấu cú pháp: *con bừa, con am, con lều...*, từ *ban* trong kết cấu *ban trống canh* [7], từ *bằng* trong kết cấu *ngọt bằng mít, mát bằng dứa...* [4].

3.2. Khảo sát các nguồn ngữ liệu hữu quan, chúng tôi thu được kết quả về lượng từ cổ trong đó như sau

NGUỒN KHẢO SÁT	SỐ LƯỢNG TỪ CỔ TỪ C	MIÊU TẢ TƯƠNG ĐỐI ĐỘ DÀI VĂN BẢN
<i>Phật thuyết</i>	131	Chính văn khoảng 25 trang in chữ quốc ngữ khổ 16 x 23,5 cm. (5.201 lượt từ ngữ).
Bốn bài phú đời Trần	104	15,5 trang in chữ quốc ngữ khổ 13 x 19cm
<i>Khoá hư lục</i>	149	Chính văn khoảng 48 trang in chữ quốc ngữ khổ 15 x 21 cm.
<i>Quốc âm thi tập</i>	221	254 bài thơ (208 bài 8 câu, 46 bài 4 câu)
<i>Hồng Đức quốc âm thi tập</i>	228	328 bài thơ (318 bài bát cú, 10 bài tứ tuyệt)
<i>Thơ Môm Ng. Bình Khiêm</i>	124	161 bài thơ bát cú
<i>Hoa Di dịch ngữ</i>	26	716 mục từ
<i>Chi Nam ngọc âm giải nghĩa</i>	236	Văn vần lục bát, hoặc bốn tiếng, sáu tiếng, 175 trang in chữ quốc ngữ, khổ 13 x 19cm.
<i>Việt sử diễn âm</i>	142	2332 câu lục bát.
<i>Truyện kì mạn lục</i> ¹⁵	168	10 truyện đầu tiên trong 20 truyện; 102 trang in chữ quốc ngữ khổ 16 x 23,5 cm.

Bảng này cho thấy một quang cảnh phân tán của các từ cổ trong các nguồn được khảo sát. Điều này không khó hiểu, bởi vì loại hình văn bản và phong cách ngôn ngữ không giống nhau, độ dài văn bản khác nhau, niên đại văn bản cũng phân bố rải từ khoảng thế kỉ XI, XII đến tận thế kỉ XVI.

Để đỡ phức tạp và trùng lặp cho quá trình phân tích, chúng tôi quy các từ thuộc năm trường hợp nêu trên thành ba tiểu loại: Tiểu loại A gồm những từ mà hiện nay coi như đã mất hẳn trong tiếng Việt hiện đại (trường hợp a. b.); Tiểu loại B gồm những từ chỉ còn tồn tại trong một vài kết cấu ngôn ngữ hoặc được sử dụng trong phạm vi hết sức hẹp (trường hợp c. và d.); Tiểu loại C gồm những từ vẫn còn lại đến nay, nhưng ngữ nghĩa, ngữ pháp của chúng đã đổi khác, những nghĩa cổ và đặc điểm, năng lực ngữ pháp cổ của chúng đã không còn được dùng nữa (trường hợp e.).

(Cần lưu ý rằng, ở đây, một số từ có thể “lưỡng khả” trong sắp xếp tiểu loại. Từ *cá*, từ *những*... chẳng hạn. *Những* với ý nghĩa “toàn là, chỉ là”, *cá* với ý nghĩa “to, lớn” có thể được qui vào tiểu loại B, vì trong tiếng Việt ngày nay, ta có các kết cấu *con cá*, *đũa cá*, *cá lo*, *cá nghĩ*... và đôi khi, một cách rất hạn chế, ta gặp những phát ngôn như “*những việc là việc*”, “*toàn những người*”

là người"...; nhưng cũng có thể qui chúng về tiểu loại C vì *cả* và *những* hiện nay đang tồn tại, nhưng *cả* chỉ dùng với ý nghĩa "toàn thể", *những* chỉ dùng với ý nghĩa "số nhiều", tức là nghĩa của chúng đã biến đổi so với trước đây).

Giữ nguyên các tiểu dị có thể có do quan niệm và cách phiên âm Nôm ở một số từ, (chưa quy chúng về thành một đơn vị từ trong danh sách chung), chúng tôi thu được kết quả phân loại chi tiết như sau:

Bảng 8. Các tiểu loại từ ngữ cổ trong các nguồn ngữ liệu được khảo sát

NGUỒN NGỮ LIỆU	TIỂU LOẠI A	TIỂU LOẠI B	TIỂU LOẠI C
<i>Phật thuyết...</i>	79	44	8
Bốn bài phú đời Trần	59	29	16
<i>Quốc âm thi tập</i>	123	82	16
<i>Hồng Đức quốc âm thi tập</i>	148	69	11
<i>Thơ Nôm Ng. Bình Khiêm</i>	76	44	4
<i>Hoa Di dịch ngữ</i>	14	11	1
<i>Chi Nam ngọc âm giải nghĩa</i>	163	60	13
<i>Khoá hư lục giải nghĩa</i>	90	52	7
<i>Việt sử diễn âm</i>	82	48	12
<i>Truyền kì mạn lục</i>	85	67	18

Số liệu trong bảng 8 cho thấy xu hướng chung trong tất cả các nguồn ngữ liệu được khảo sát là: số lượng các từ tiểu loại A luôn luôn lớn nhất, tiểu loại B thứ hai nhưng ít hơn nhiều so với tiểu loại A; và tiểu loại C là tiểu loại có số lượng rất nhỏ so với hai tiểu loại B và A.

3.3. Từ kết quả khảo sát trên đây, đem tổ hợp thành một danh sách chung, chúng tôi thu được tất cả là 949 từ khác nhau, phân bố ở tiểu loại A: 640 từ, tiểu loại B: 247 từ, tiểu loại C: 62 từ.

Trên thực tế, không phải chỉ có sự không đồng đều về lượng giữa các tiểu loại từ cổ được khảo sát. Năng lực hoạt động, sự phân bố (có mặt) của mỗi từ trong các nguồn ngữ liệu được khảo sát cũng không đồng đều: từ thì có trong nhiều nguồn, từ thì có trong ít, thậm chí chỉ trong một nguồn. Số liệu cụ thể như sau:

Bảng 9. Phân bố của các tiêu loại từ cổ trong các mẫu nghiên cứu

	TIÊU LOẠI A	TIÊU LOẠI B	TIÊU LOẠI C	TỔNG A + B + C
Số từ có trong 10 nguồn	0	1	0	1
Số từ có trong 09 nguồn	5	2	1	8
Số từ có trong 08 nguồn	3	1	0	4
Số từ có trong 07 nguồn	2	6	0	8
Số từ có trong 06 nguồn	7	5	3	15
Số từ có trong 05 nguồn	10	10	3	23
Số từ có trong 04 nguồn	11	14	2	27
Số từ có trong 03 nguồn	15	23	1	39
Số từ có trong 02 nguồn	70	40	5	115
Số từ có trong 01 nguồn	517	145	47	709
TỔNG HỢP 10 NGUỒN	640	247	62	949

Danh sách cụ thể những từ có mặt trong ba nguồn được khảo sát trở lên như sau:

Tiêu loại A.

5 từ có trong 9 nguồn: *chín (chi), chung, màng/ mǎng (nghe), mưa (đùng/ chớ), thừa.*

3 từ có trong 8 nguồn: *khá (đáng, nên), nhẫn (đến/ tới), no (đủ).*

2 từ có trong 7 nguồn: *hoà (và/ cả), sá (nên)*

7 từ có trong 6 nguồn: *cóc/ cốc (biết/ hiểu), dầu (tuỳ thích), ghê (nhiều), khùng (chịu), la đá (đá), nèo (khi/ lúc), tua (nên).*

10 từ có trong 5 nguồn: *bui (duy), đòi (nhiều), ghẽ (rẽ/ chia), ghín (cẩn thận), lọ (huống hồ), óc / óc (gọi /*

kêu), phô (các), rập (giúp), tượng (tựa như), xây/ xáy (chợt).

11 từ có trong 4 nguồn: *áy (héo), khăn khăn (có một lòng), hợp (nên), loạn đần (bừa bãi, càn bậy), mỗ (tôi/ ta), mỗ giáp (tôi), phen (so/ bì), song le, song viết, thay tháy (tất cả), vay! (ngữ khí từ).*

15 từ có trong 3 nguồn: *bô (nhiều), chiếng (nhú ra), chốc mòng, đăm (đem, ham), đằm hằm, dâm (đen), dǎng dôi, díp/ điép (lúc trước), đoạn, hằm hằm (mạnh bạo), khá (có thể), nuôi (mày; từ xưng gọi), rân (tiếng -), ru (ngữ khí từ), thúc thúc (lựa thô).*

Tiêu loại B.

1 từ có trong 10 nguồn: *cả (lớn)*

2 từ có trong 9 nguồn: *khôn (không thể), đòi (theo).*

từ có trong 8 nguồn: *lệ (sợ/ ngại)*.

6 từ có trong 7 nguồn: *dấu (yêu), đấng (bậc), há (phó từ), hằng (thường), hầu (ngõ hầu/ nhằm), luống (xây ra nhiều lần)*.

5 từ có trong 6 nguồn: *âu (lo), chiêu (bên trái), cuu (mang, hạp lại), dăm (bên phải), rày/ rầy (nay)*.

10 từ có trong 5 nguồn: *chác (mua/ đổi), chãng (chẳng), đường (loại, thứ), gìn (giữ), kham (chịu được), một (chi, toàn), ngõ (nhằm), rồi (rối), tày (bằng), tây (riêng)*.

14 từ có trong 4 nguồn: *áng (đám), chừ (bây giờ), chước (muru), dái (sợ), hay (biết/ có thể), khó (nghèo), kíp (vội), lường (lo liệu), mẩy (vớ), nhặt (mau), tác (tuổi), thông (đủ), và (và lại), vì (nê, giúp)*.

23 từ có trong 3 nguồn: *cái (mẹ), chầy (lâu), chiền (chùa), dấy (nổi/ vùng lên), dôi (noi), đương (nên), gẫm (ngẫm), han (hỏi), lăm (lăm lăm), hổ (xấu hổ), lộn (trộn), mọn (nhỏ), ngợi (khen), nà (bao -), rẻ (khinh), rờ rờ (sáng), sốt (nóng), thầy (tát cà), thể (- gian), thốt (nói), vọt (roi), xâm (lấn), xống (váy)*.

Tiểu loại C

1 từ có trong 9 nguồn: *bằng (từ so sánh, cách dùng cổ)*.

3 từ có trong 6 nguồn: *ban (buổi, khi), đõ (ở/ nghi), nữa (hơn)*.

3 từ có trong 5 nguồn: *cái (= con), lâu/ lâu lâu (rõ/ hoàn toàn), những (toàn là)*.

2 từ có trong 4 nguồn: *rằng (nói), thức (màu, sắc)*.

1 từ có trong 3 nguồn: *con (= cái)*.

Để có thể có được một hình dung rõ hơn về diện mạo và sự hoạt động của một số từ ngữ cổ trong giai đoạn đang xét, chúng ta hãy quan sát hai đoạn trích ngắn sau đây. (Dấu chấm, dấu phẩy là do người phiên sang chữ quốc ngữ thêm vào cho dễ hiểu. Những từ ngữ cổ được in nghiêng là do chúng tôi thực hiện).

Một đoạn trong *Phật thuyết đại báo phụ mẫu ân trọng kinh*:

“Bụt bảo A Nan:

Tao xem chúng sinh tuy được làm loài người, trông nét chẳng quen, tối mờ, chẳng lo *áng nạ* có ân đạo cả, chẳng ra lòng cung kính, rời ân, bội đạo, chẳng có lòng *dấu xót*, chẳng thảo, chẳng có nghĩa. (...) Nuôi nấng dường ấy vậy cho nên người. Đến con cả nên người, lại làm bê chẳng thảo, *áng nạ thốt cùng* thừa trả *án úi* mà lét mắt, mà trùng *con sâu*, dối cùng chú bác, đánh mắng *anh tam*, nhục để họ hàng, chẳng có lễ nghĩa, chẳng đòi thầy dạy, *áng nạ* dạy bảo *trộn* chẳng đòi *lọt*. *Anh tam* mà *thốt cùng* *đương* cùng *gượng* đòi, vào ra đi lại chẳng thừa người cả, lời nét bừa bộn, *dầu* lòng làm việc, *áng nạ* dạy trách, chú bác *chê* đàn. Con trẻ *dấu xót*, *kẻ cả* che hộ, dần dần nên cả, *cưỡng* *lượng* chẳng hoà, chẳng *dái*, lời nét ngược dữ, cùng lại ra lòng hờn giận, rời bao nhiêu bạn lành, làm bạn gần dữ, *duợt* ã nên tính, mới làm những kẻ đại, *dối* người *đõ thốt*, trốn đi *nước khác*, trái lời *áng nạ*, xa nhà đi *nước khác*, hoặc là nhân buôn bán, hoặc vì đi *chác*, lẫn lóc luân hồi mới nên *thá* vợ chồng, nhân vậy mới *đõ ở* lâu chẳng về nhà...” [15, 154-155].

Đoạn đầu tiên của truyện thứ nhất, quyền một trong *Truyện kì mạn lục giải âm*:

“Quan Thừa chi họ Hồ, tên là Tông Thốc, hay *chung* thơ, càng dài *chung* lời dạy dỗ, lời trêu giễu. Rốt đời nhà Trần, vâng mệnh sang sứ nước Bắc. Qua dưới miếu Hạng Vương, *bày* thơ rằng. (...) *Bày* thơ rồi, giống vọt ngựa lại nghỉ điểm khách. Rượu *hầu* say, muốn ngủ. Thấy một người đến trước, có lời rằng: “Vâng chiếu chỉ vua ta, phiên người đến *nói* khó”. Họ Hồ bèn vội vàng khép nép sửa sang. *Thừa* người ấy bèn dẫn đi mé tả. Đến thời thấy điện nhà cao *cá*, quan theo châu sắp hàng. Hạng Vương trước đã ngồi ở đấy. Bên đặt giường lưu li. *Áp* người Hồ đến tiệc. Hỏi rằng: “Câu thơ đề ban ngày, người sao thấy *chê* lắm vậy *vay*? *Thừa* câu rằng “Nhất bại hữu thiên, vong trạch tả. Trùng lai vô đại đáo Giang Đông” thời thực *phải* vậy. Đến *chung* câu “Kinh dinh ngũ đài thành hà sự. Tiêu đắc khu khu táng Lỗ Công” chẳng bèn *chê* bàn *thất* đáng *ru*? *Chung* nhà Hán có muôn cỗ xe vậy, ta cũng có muôn cỗ xe vậy. Ta chẳng *hay* diệt nhà Hán, nhà Hán sao lại *hay* phong tước cho ta vậy *vay*...” [9, 29-30].

4. Đến đây, chúng tôi thấy có thể rút ra một vài nhận xét bước đầu như sau

4.1. Các chứng tích của từ vựng giai đoạn tiếng Việt cổ thể hiện rằng, vào thời đó, có rất nhiều đặc điểm cổ (so với hiện nay) còn được lưu giữ, nhất là các đặc điểm về hình thái của từ nói riêng và đơn vị từ vựng nói chung. Điều này cần được khảo sát và phân tích kĩ hơn nữa để phục

vụ cho nhiều mục đích nghiên cứu, trong đó có mục đích làm sáng tỏ hơn quá trình đơn tiết hoá triệt để, quá trình cấu tạo từ của tiếng Việt cũng như các đặc điểm loại hình của nó.

4.2. Những kết quả phân tích về từ Hán Việt và tương quan Hán Việt - phi Hán Việt trong các nguồn ngữ liệu được khảo sát sẽ có tiếng nói của chúng trong khi nghiên cứu về quá trình hình thành, xây dựng và phát triển ngôn ngữ văn hoá, văn học của dân tộc. Các bậc tiền nhân đã không ngần ngại vay mượn, sử dụng các yếu tố ngoại lai, biến “ngoại lực” thành “nội lực” để làm phong phú, giàu có thêm cho ngôn ngữ của mình. Tuy nhiên, trong phạm vi ngữ liệu quan sát được ở đây, các từ ngữ Hán Việt đã không “lấn lướt” các từ ngữ phi Hán Việt cả về số lượng lẫn năng lực hoạt động, chiếm lĩnh độ dài văn bản.

4.3. Ngót một nghìn từ cổ được nhận diện tuy chưa phải toàn bộ từ cổ của tiếng Việt thuộc giai đoạn đang xét, nhưng đó là những cứ liệu hiện thực, đáng tin cậy, hoàn toàn có thể kiểm chứng được để chúng ta có thêm cơ sở hình dung về một bình diện của từ vựng tiếng Việt cổ. Trong số các từ cổ đó, có đại diện của cả thực từ lẫn hư từ và của các từ loại. Nghiên cứu các từ cổ một cách cẩn thận, chắc chắn sẽ đem đến cho chúng ta những căn cứ đáng tin cậy để phân tích, đánh giá về mặt phát triển và biến đổi của từ vựng nói riêng, góp phần phân tích, đánh giá về sự phát triển và biến đổi của tiếng Việt nói chung.

CHÚ THÍCH

¹ Đây là văn bản viết bằng chữ Nôm bản dịch sang tiếng Việt (từ một bản kinh viết bằng chữ Hán) được Nguyễn Tài Căn [13] đặt giả thuyết niên đại vào khoảng thế kỉ XI, Nguyễn Quang Hồng [14] cho rằng văn bản ra đời khoảng đầu thế kỉ XII. Thế kỉ XI, XII thuộc giai đoạn tiếng Việt tiền cổ [12]; nhưng vì Phật thuyết là văn bản Nôm duy nhất thuộc giai đoạn này còn sót lại (tính đến nay) nên chúng tôi đặt chung vào với các nguồn thuộc giai đoạn tiếng Việt cổ để cùng khảo sát.

² Tuy nhiên, điều này không có nghĩa rằng trong mọi văn bản Nôm, ở bất kì niên đại nào, cứ có chữ Nôm được ghi bằng hai mã chữ thì chúng ta kết luận hai mã chữ đó biểu hiện hai thành tố của tổ hợp phụ âm. Nhưng đó là vấn đề về cách thức và phương pháp phiên âm chữ Nôm, đọc chữ Nôm, không trực tiếp liên quan đến những phân tích về mặt từ vựng, nên ở đây chúng tôi không thảo luận.

³ Việc Trần Cương Trung (thế kỉ XIII) trong Sứ Giao châu tập dùng hai chữ Hán để “phiên âm” một từ Việt khi “dịch” Hán - Việt, được Lê Quý Đôn ghi trong Kiến văn tiểu lục, là một minh chứng có rất có giá trị. Cụ thể là: Từ Hán đọc theo âm Hán Việt ngày nay là thiên, nghĩa Việt là trời được “phiên âm” là BỆT - LÔI (= blời trong tiếng Việt cổ). Từ Hán đọc theo âm Hán Việt ngày nay là nhật, nghĩa Việt là mặt trời được phiên âm là MAT BỆT - LÔI (= mặt blời trong tiếng Việt cổ). Từ Hán đọc theo âm Hán Việt ngày nay là nguyệt, nghĩa Việt là (mặt) trăng được “phiên âm” là MAT LĂNG (= mặt blăng trong tiếng Việt cổ).

⁴ Trên đây, chúng tôi phân biệt hai kiểu từ, một kiểu còn giữ cấu trúc CvCVC, một kiểu còn giữ tổ hợp phụ âm CC trong cấu trúc CCVC, mà kiểu nào cũng đều được ghi bằng chữ Nôm gồm hai mã chữ Hán cả. Về căn bản, mặt văn tự cho thấy: cấu trúc CvCVC được ghi bằng hai mã chữ tách rời nhau, còn cấu trúc CCVC được ghi bằng hai mã chữ hội nhập làm một.

Sự phân biệt này là khả tín, nhưng phải lúc nào cũng có thể hoàn toàn quyết định một cách tuyệt đối minh xác, bởi vì [x.14]:

a) Có khi một mã chữ (như mã chữ Cá chẳng hạn), ở chỗ này được dùng để biểu thị một thành tố trong tổ hợp CC, những chỗ khác lại chỉ được dùng như một kí tự phụ để báo hiệu sự “hiệu chỉnh” âm đọc, phải “đọc trệch đi”, vì thực chất, chữ Nôm, nhất là chữ Nôm trong thời kì khai mở, không có một điển chế hoàn toàn chặt chẽ. Thí dụ: các trường hợp như CÁ+LUNG (= trông), CÁ+LAI (= trái)... thì CÁ được dùng để biểu thị thành tố [k-] của tổ hợp [kl], nhưng những trường hợp như CAO - CÁ (= gào), ĐAI - CÁ (= dạy), MỘC - CÁ (= móc)... thì CÁ được dùng để báo hiệu sự hiệu chỉnh âm đọc.

b) Có khi một từ, ở chỗ này được ghi bằng hai mã chữ tách rời, chỗ khác lại được ghi bằng hai mã chữ nhập một, chỗ khác nữa lại có thể được ghi chỉ bằng một mã chữ. Thí dụ: Từ lồi, khi thì được ghi bằng hai mã chữ tách rời MA - LÊ, khi thì được ghi bằng hai mã chữ nhập một MA + LÊ, khi thì được ghi chỉ bằng một mã chữ LÊ. Từ trông, khi thì được ghi bằng hai mã chữ tách rời CÁ-LUNG, khi thì được ghi bằng hai mã chữ nhập một LONG+ CỤ, khi thì được ghi chỉ bằng một mã chữ LUNG.

c) Đôi khi, có những từ được ghi bằng hai mã chữ tách rời nhưng không đọc theo cấu trúc CvCVC, mà phải đọc theo cấu trúc CCVC và ngược lại. Các trường hợp ghi theo kiểu CỤ-LUNG (đọc là klông = trông), MA LÊ (đọc là mlời = lồi), MA-LĂN (đọc là mlón = lớn), MA-LẢ (đọc là mlờ = lờ)... là như vậy.

“Có thể nghĩ rằng tuy thời bấy giờ tiếng Việt vẫn hiện hữu cấu trúc Cv-CVC, song ở không ít từ đã bắt đầu có hiện tượng phát âm lưỡng khả với cả cấu trúc CCVC, không ổn định, dễ rồi chuyển dần sang hình thức một âm tiết đơn nhất... Đây là tình trạng điển hình của một ngôn ngữ đang chuyển biến mạnh từ loại hình ngôn ngữ đơn lập cận âm tiết tính sang loại hình đơn lập âm tiết tính” [14, 141].

Những điều trên đây thể hiện rằng: quá trình giản hoá cấu trúc CvCVC của từ và giản hoá cấu trúc CCVC của âm tiết tiếng Việt (đơn vị hình thái học căn bản của nó) trong thời kì chúng ta đang đề cập ở đây, vẫn đang tiếp tục diễn ra, chưa kết thúc, chưa đưa đến cấu trúc giản hoá triệt để như trong tiếng Việt ngày nay.

⁵ Về mặt thống kê, cần phải thấy rằng, trên cùng một mẫu khảo sát (ở mẫu khảo sát này cũng như các mẫu khác), số liệu tuyệt đối thu được giữa người này với người khác có thể xê xích nhau chút ít, do những phức tạp của việc phân định ranh giới từ và xác định bản chất ngữ pháp của từ ở những trường hợp giáp ranh... luôn luôn gây rất nhiều khó khăn cho người thực hiện. Tuy nhiên, độ xê xích, sai biệt đó (nếu có), thường chỉ là, và phải được bảo đảm chỉ là sai số cho phép để bảo đảm độ tin cậy của kết quả thống kê, không ảnh hưởng đến các ước lượng, miêu tả và phân tích, đánh giá.

⁶ Số liệu này, chúng tôi kiểm đếm được trên toàn văn bản phiên sang chữ quốc ngữ trong [15] của tác giả Hoàng Thị Ngọc, không phải là rút ra được từ bảng "Sách dẫn" trong công trình đó. Ở đây, chúng tôi chỉ điều chỉnh cách đọc của một từ (song ngữ > song viết - tr.154), và không tính vài ba từ/chữ bị rách mất trong nguyên bản chữ Nôm, cũng không tính những chữ được tác giả đoán định, bổ sung, đặt trong ngoặc vuông, không tính khoảng mười đến mười lăm chữ Nôm nữa chưa được giải quyết dứt khoát về mặt âm đọc nôm chưa đưa vào bản phiên sang chữ quốc ngữ (theo trao đổi riêng với chị). Tuy nhiên, số từ/chữ này quá nhỏ so với toàn bộ độ dài của văn bản, không thể làm ảnh hưởng đến những đánh giá ở đây. Mặt khác, chúng tôi cũng xin lưu ý rằng bản phiên chuyển sang chữ quốc ngữ hiện đại không phân ánh các từ có cấu trúc CvCVC. Nhưng dù có cộng thêm khoảng trên 100 từ có cấu trúc CvCVC [theo 14, 139] vào số các đơn vị từ vựng đa tiết ở đây thì cũng không làm thay đổi được đánh giá về tương quan toàn cục giữa số đơn vị từ vựng đơn tiết với số

đơn vị đa tiết ở văn bản được khảo sát. Bên cạnh đó, từ góc độ từ vựng học, chúng tôi kiểm đếm các đơn vị từ vựng, chứ không chỉ kiểm đếm các từ, càng không phải là kiểm đếm các chữ (Nôm). Vì vậy, mỗi đơn vị từ vựng đa tiết chẳng hạn như họ hàng, láng giềng, tui hồ... sẽ được coi là một đơn vị thống kê.

⁷ Năm 1980, trong một nghiên cứu của mình, chúng tôi công bố kết quả thống kê trong Quốc âm thi tập có 2.235 từ, gồm 1590 từ đơn và 645 từ ghép (phức). Kết quả kiểm đếm ở đây có được hiệu chỉnh lại vì không tính các tên riêng và có phân định lại ranh giới của một số từ ngữ cụ thể [17].

⁸ Các số liệu về Hồng Đức quốc âm thi tập được kiểm đếm, tính toán dựa trên bảng từ do Phạm Xuân Chiến thực hiện (với sự hướng dẫn của chúng tôi), sau khi chúng tôi đã xử lí lại toàn bộ.

⁹ Chúng tôi lập danh sách từ ngữ và kiểm đếm căn cứ vào các từ ngữ Việt được các tác giả của cuốn dịch ngữ sử dụng để dịch 716 mục từ Hán (Hoa) trong cuốn dịch ngữ của họ [1].

¹⁰ Thực ra, ở đây có một vấn đề cần được lưu ý là giữa Quốc âm thi tập với Thơ Nôm Nguyễn Bình Khiêm có 27 bài thơ trùng nhau hoặc gần trùng nhau, cần phải được tiếp tục nghiên cứu để minh định tác giả. Điều này rất quan trọng trong khi nghiên cứu văn học sử hoặc nghiên cứu tác gia, tác phẩm; nhưng đối với việc nghiên cứu từ vựng thì dù của tác giả nào, 27 bài thơ đó cũng vẫn là sản phẩm ngôn ngữ của thế kỉ XV-XVI, niên đại của tiếng Việt cổ.

¹¹ Trường hợp văn bản Phật thuyết, như chúng tôi đã đề cập trong chú thích ⁽²³⁾ bên trên, nếu truy tìm lại các từ có cấu trúc CvCVC để cộng vào số các đơn vị đa tiết thì tương quan đơn tiết - đa tiết ở đây vẫn không thay đổi.

¹² Xem chú thích ⁽⁶⁾ bên trên.

¹³ Số liệu đã được hiệu chỉnh lại so với kết quả trình bày trong [17] của chúng

tôi vì ở đây chúng tôi không tính các tên riêng và có phân định lại ranh giới của một số trường hợp cụ thể.

¹⁴ *Thí dụ như:* buông, buông, búa, cái, cần, che, chuộc, dời, dưng, đủ, đỡ, đúng, đĩa, giềng, giường, hòm, hẹp, hè, kim, kép, khoe, khéo lia, lừa, mảy, mạng, mong, ngờ, ngược, phen (lần), quen, say, tiếc, tựa, thêm, việc, vua, xe, xét, xưa... dưng, dao, dài, đáng, gân, gươm, gồm, kém, khăn, ngờ, rèm, sen, sức, thầy, thả, tim, và, vái, ven, vuông...

¹⁵ *Chúng tôi sử dụng bảng từ của Phạm Đức Hiếu trong khoá luận tốt nghiệp Trường đại học khoa học xã hội và Nhân văn, Đại học Quốc gia Hà Nội, 2007; nhưng đã đối chiếu, kiểm tra và xử lý lại để có số liệu này.*

NGUỒN NGŨ LIỆU

1. An Nam dịch ngữ, Hoa Di dịch ngữ, *Vương Lộc Giới thiệu và chú giải*, Nxb Đà Nẵng - Trung tâm từ điển học, 1995.
 2. Chi Nam ngọc âm giải nghĩa, *Trần Xuân Ngọc Lan phiên âm, chú giải*, Nxb KHXH, H., 1985.
 3. Cư trần lạc đạo phú, *Đặc thú lâm tuyền thành đạo ca*, Hoa yên tự phú, Phú dạy con, *trong Chữ Nôm (nguồn gốc - cấu tạo - diễn biến)*, Đào Duy Anh Nxb KHXH, H., 1975.
 4. Hồng Đức quốc âm thi tập, *Phạm Trọng Diễm - Bùi Văn Nguyên phiên âm, chú giải, giới thiệu*, Nxb Văn hoá, 1962, Nxb Văn học, 1982.
 5. Phật thuyết phụ mẫu đại báo ân trọng kinh, *Hoàng Thị Ngọc phiên âm, chú giải, trong Chữ Nôm và tiếng Việt qua bản giải âm "Phật thuyết phụ mẫu đại báo ân trọng kinh"*, Nxb KHXH, H., 2002.
 6. Quốc âm thi tập, *Nguyễn Trãi, trong Nguyễn Trãi toàn tập, Đào Duy Anh phiên âm, khảo chú*, Nxb KHXH, H., 1976.
 7. Thiên tông khoá hư ngữ lục, *Trần Thái tông, (Trần Trọng Dương khảo cứu, phiên chú)*, Nxb Văn học, H., 2009.
 8. Thơ Nôm Nguyễn Bình Khiêm, *trong Thơ văn Nguyễn Bình Khiêm, Nxb Văn học, H., 1983.*
 9. Truyền kì mạn lục giải âm, *Nguyễn Quang Hồng phiên âm, chú giải*, Nxb KHXH, H., 2001.
 10. Việt sử diễn âm, *(Nguyễn Tá Nhi sưu tầm, giới thiệu, biên dịch)*, Nxb Văn hoá thông tin, 1997.
- #### TÀI LIỆU THAM KHẢO
11. *Nguyễn Tài Cán*, Giáo trình lịch sử ngữ âm tiếng Việt, Nxb GD, H., 1995.
 12. *Nguyễn Tài Cán*, Một số chứng tích về ngôn ngữ, văn tự và văn hoá, Nxb ĐHQG, H., 2001.
 13. *Nguyễn Tài Cán*, Chữ Nôm đã hiện diện vào thời kì quốc đô dời ra Thăng Long, *Văn hoá Nghệ An*, Số 172 - 173, 2010.
 14. *Nguyễn Quang Hồng*, Khái luận văn tự học chữ Nôm, Nxb GD, H., 2008.
 15. *Hoàng Thị Ngọc*, Chữ Nôm và tiếng Việt qua bản giải âm "Phật thuyết phụ mẫu đại báo ân trọng kinh", Nxb KHXH, H., 2002.
 16. *Shimizu Masaaki*, Khảo sát sơ lược về cấu trúc âm tiết tiếng Việt vào thế kỉ XIV - XV qua hai cứ liệu chữ Nôm, *(Kỷ yếu Hội thảo quốc tế lần thứ nhất về Việt Nam học lần thứ nhất, H., 1998, tr. 252-265.*
 17. *Vũ Đức Nghiệu - Nguyễn Tài Cán*, Một vài nhận xét bước đầu về ngôn ngữ trong thơ Nguyễn Trãi (qua số liệu thống kê), *T/c Ngôn ngữ*, Số 3, 1980.
 18. *Vũ Đức Nghiệu*, Hư từ tiếng Việt thế kỉ XV trong "Quốc âm thi tập" và "Hồng Đức quốc âm thi tập", *T/c Ngôn ngữ*, Số 12, 2006.
 19. *Stankievich N.V.*, Vài nhận xét về các hư từ tiếng Việt thế kỉ 16 (Tư liệu rút từ Truyền kì mạn lục giải âm), *T/c Ngôn ngữ*, Số 9, 2006.