

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC KHOA HỌC XÃ HỘI VÀ NHÂN VĂN

Họ và tên tác giả luận án

Li Lijuan

TÊN ĐỀ TÀI LUẬN ÁN

**KHẢO SÁT ĐẶC ĐIỂM TỪ VỰNG TRÊN “西贡解放日报”-
CHUYÊN TRANG CỦA BÁO SÀI GÒN GIẢI PHÓNG ĐIỆN TỬ
(ĐỐI CHIẾU VỚI TIẾNG HÁN PHỔ THÔNG VÀ TIẾNG VIỆT)**

Chuyên ngành: Ngôn ngữ học so sánh – đối chiếu

Mã số: 9229020.03

TÓM TẮT LUẬN ÁN TIẾN SĨ NGÔN NGỮ HỌC

Hà Nội - 2026

Công trình được hoàn thành tại: Trường Đại học Khoa học Xã hội và Nhân văn,
Đại học Quốc gia Hà Nội.

Người hướng dẫn khoa học: PGS.TS. Nguyễn Hoàng Anh

TS. Võ Thị Minh Hà

Phản biện: GS. TS. Nguyễn Văn Khang

Viện Ngôn ngữ học, Viện Hàn lâm Khoa học Xã hội Việt Nam

Phản biện: . PGS. TS. Cẩm Tú Tài

Trường Ngoại ngữ - Du lịch, Đại học Công nghiệp Hà Nội

Phản biện: PGS. TS. Trần Trọng Dương

Trường Khoa học Liên ngành và Nghệ thuật, ĐHQGHN

Luận án sẽ được bảo vệ trước Hội đồng cấp Đại học Quốc gia chấm luận án
tiến sĩ họp tại phòng 302, nhà E, trường Đại học Khoa học Xã hội và Nhân văn
số 336 Nguyễn Trãi, Thanh Xuân, Hà Nội.

Vào hồi 14 giờ ngày 9 tháng 7 năm 2026

Có thể tìm hiểu luận án tại:

- Thư viện Quốc gia Việt Nam
- Trung tâm Thông tin - Thư viện, Đại học Quốc gia Hà Nội

MỞ ĐẦU

1. Lý do chọn đề tài

Trong bối cảnh toàn cầu hóa, tiếng Hoa tại Việt Nam không chỉ là phương tiện giao tiếp, mà còn đóng vai trò gắn kết cộng đồng và bảo tồn di sản văn hóa của người Hoa nơi đây. Quá trình vận động của tiếng Hoa trong các không gian ngôn ngữ mới luôn thể hiện sự song hành giữa tính bảo tồn (những đặc điểm cốt lõi của tiếng Hán) và tính biến đổi (sự thích ứng với điều kiện xã hội - ngôn ngữ sở tại). Tại Việt Nam, quy luật này càng trở nên đặc biệt do tiếng Việt có lịch sử vay mượn và tiếp xúc lâu dài với tiếng Hán, khiến cho biến thể tiếng Hoa tại đây mang những đặc điểm lai ghép độc đáo, phản ánh mối quan hệ văn hóa - lịch sử sâu sắc giữa hai quốc gia.

Trong nghiên cứu ngôn ngữ học, đặc biệt là ngôn ngữ học đôi chiếu, báo chí được xem là nguồn ngữ liệu giá trị, phản ánh chân thực đời sống ngôn ngữ - xã hội của cộng đồng. Báo Sài Gòn Giải Phóng (华文西贡解放日报) là trường hợp tiêu biểu: đây là tờ báo tiếng Hoa duy nhất tại Việt Nam, cơ quan ngôn luận của Đảng bộ, chính quyền và nhân dân Thành phố Hồ Chí Minh, phục vụ cộng đồng người Hoa. Đặc biệt, chuyên trang điện tử của báo không đơn thuần là bản dịch từ tiếng Việt mà được biên tập độc lập với định hướng nội dung riêng, hình thành một lớp từ vựng đặc hữu phản ánh đời sống văn hóa - xã hội của người Hoa Việt Nam (ví dụ qua các chuyên mục như “Độc giả – Từ thiện”, “Văn hóa – Giải trí”).

Chính sự độc lập này dẫn đến những khác biệt đáng chú ý trên bình diện từ vựng học. Hệ thống từ vựng trên báo không chỉ bao gồm lớp từ cơ bản của tiếng Hán phổ thông mà còn tích hợp nhiều đơn vị ngôn ngữ mang dấu ấn bản địa, tạo thành một dạng thức từ vựng đa dạng và lai ghép.

Vì vậy, nghiên cứu từ vựng trên tờ báo có hai ý nghĩa quan trọng:

Làm rõ đặc trưng ngôn ngữ của tiếng Hán trong bối cảnh tiếp xúc với tiếng Việt; Nhận diện những nét độc đáo của tiếng Hoa tại Việt Nam trong mối tương quan so sánh với các biến thể tiếng Hoa khác trong khu vực và trên thế giới.

Từ những lý do trên, đề tài được lựa chọn nhằm góp phần làm phong phú thêm các nghiên cứu về tiếng Hoa ngoài Trung Quốc đại lục nói chung và biến thể tiếng Hoa tại Việt Nam nói riêng.

2. Mục đích nghiên cứu

Mục đích của luận án là thông qua việc phân tích hệ thống từ vựng trên Chuyên trang *Báo Sài Gòn Giải Phóng* điện tử, làm rõ hiện tượng chuyển di ngữ nghĩa và biến đổi từ vựng trong môi trường văn hoá – xã hội Việt Nam. Kết quả nghiên cứu nhằm góp phần nhận diện cơ chế tiếp xúc ngôn ngữ Hán – Hoa, Hán – Việt trong đời sống báo chí và chỉ ra những đặc thù của tiếng Hán tại Việt Nam.

3. Nhiệm vụ nghiên cứu

- Hệ thống hoá và lựa chọn các lý thuyết nền tảng về từ vựng học, tiếp xúc ngôn ngữ và biến thể ngôn ngữ, đối chiếu ngôn ngữ.

- Thu thập, số hoá và xử lý để hình thành khối liệu báo SGGP, đảm bảo tính đại diện và khoa học.

- Rà soát, nhận diện và phân loại các lớp từ vựng đặc thù, có sự khác biệt so với tiếng Hán phổ thông và có liên quan đến từ Hán Việt tương ứng.

- Miêu tả một cách hệ thống các đặc điểm từ vựng đã được nhận diện trên các phương diện nghĩa, kết hợp từ, sắc thái biểu cảm và tần suất sử dụng; đối chiếu với tiếng Hán phổ thông và từ Hán Việt.

- Phân tích nguyên nhân hình thành, tập trung làm rõ vai trò của các yếu tố xã hội, văn hoá và lịch sử tiếp xúc Việt – Hán.

4. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu của luận án là hệ thống từ vựng trên chuyên trang Báo Sài Gòn Giải Phóng điện tử, tập trung phân tích đặc điểm ngữ nghĩa, hình thái, khả năng kết hợp và tần suất sử dụng, cùng các yếu tố xã hội - văn hóa chi phối. Phạm vi nghiên cứu giới hạn ở 9 chuyên mục tiêu biểu (时政, 社会, 法律, 经济, 华人动态, 教育, 体育, 文娱, 读者-慈善, 旅游), khảo sát ngữ liệu từ tháng 3/2017 đến tháng 3/2025, và chỉ tập trung vào bình diện từ vựng - ngữ nghĩa cùng hệ quả về cách dùng, không mở rộng sang phân tích cú pháp hay văn phong.

5. Phương pháp nghiên cứu

Luận án sử dụng các phương pháp và thủ pháp nghiên cứu gồm phương pháp ngôn ngữ học khối liệu, phương pháp so sánh, đối chiếu và thủ pháp miêu tả, phân tích.

6. Ý nghĩa khoa học và thực tiễn của luận án

Về lý thuyết, nghiên cứu góp phần làm phong phú thêm hiểu biết trong lĩnh vực từ vựng học, ngôn ngữ học đối chiếu và ngôn ngữ học xã hội thông qua việc phân tích đặc điểm tiếp xúc Hán – Hoa, Hán - Việt và quá trình thích ứng của tiếng Hán trong môi trường phi bản ngữ. Về thực tiễn, kết quả nghiên cứu cung cấp cơ sở cho công tác biên tập và sử dụng ngôn ngữ trên báo chí Hoa văn tại Việt Nam, đồng thời đóng góp vào các nghiên cứu so sánh về tiếng Hoa trong khu vực Đông Nam Á và toàn cầu.

7. Bố cục của luận án

Ngoài phần mở đầu, kết luận, danh mục tài liệu tham khảo, phụ lục, nội dung luận án được kết cấu thành 4 chương như sau:

Chương 1: Tổng quan tình hình nghiên cứu và cơ sở lý thuyết của đề tài.

Chương 2: Diện mạo từ vựng trên Chuyên trang Báo *Sài Gòn Giải Phóng* điện tử.

Chương 3: Đối chiếu từ vựng trên Chuyên trang Báo *Sài Gòn Giải Phóng* với tiếng Hán phổ thông.

Chương 4: Đối chiếu từ vựng trên Chuyên trang Báo *Sài Gòn Giải Phóng* với lớp từ Hán Việt

CHƯƠNG 1. TỔNG QUAN TÌNH HÌNH NGHIÊN CỨU VÀ CƠ SỞ LÝ THUYẾT

1.1. Tổng quan tình hình nghiên cứu

1.1.1. Hướng tiếp cận từ góc độ truyền thông báo chí

Sài Gòn được xem là cái nôi của báo chí Việt Nam, và báo *Sài Gòn Giải Phóng* (Hoa văn) được xác định là tờ báo Hoa ngữ duy nhất tại Việt Nam, có vị thế đặc biệt trong hệ thống báo chí và cộng đồng người Hoa. Các nghiên cứu trong nước Việt Nam chủ yếu mang tính mô tả lịch sử, trong khi một số học giả Trung Quốc đã phát hiện những

khác biệt về từ ngữ và ngữ pháp trên tờ báo này so với tiếng Hán phổ thông.

1.1.2. Hướng tiếp cận từ góc độ so sánh, đối chiếu từ vựng Hán -Hoa, Hán - Việt

Từ góc độ Hán - Hoa, giới nghiên cứu Trung Quốc thống nhất xem tiếng Hoa là biến thể của tiếng Hán, được hình thành trong cộng đồng Hoa kiều để thích ứng với môi trường văn hóa - xã hội bản địa. Các nghiên cứu so sánh từ vựng Hán - Hoa tại Đông Nam Á thường sử dụng báo chí Hoa văn làm ngữ liệu, chỉ ra những khác biệt chủ yếu về từ vựng và ngữ pháp. Từ góc độ Hán - Việt, các nghiên cứu tập trung phân tích từ Hán Việt trên các phương diện ngữ nghĩa, ngữ âm, ngữ pháp và hiện tượng “Việt hóa” trong cấu trúc từ vựng, đồng thời khảo sát mức độ tương đồng và dị biệt so với từ vựng tiếng Hán hiện đại.

1.1.3. Một vài nhận xét về tình hình nghiên cứu

Các nghiên cứu hiện có về báo chí Hoa ngữ tại Việt Nam chủ yếu tập trung vào khía cạnh truyền thông, trong khi nghiên cứu ngôn ngữ học, đặc biệt về từ vựng, còn nhiều khoảng trống đáng kể. Đáng chú ý, chưa có công trình nào hệ thống hóa từ vựng tiếng Hoa trên báo Sài Gòn Giải Phóng thông qua đối chiếu song song với cả tiếng Hán phổ thông và tiếng Việt. Phần lớn nghiên cứu trước đây chỉ giới hạn ở so sánh Hán-Việt, bỏ qua trục Hoa-Việt trong bối cảnh đa ngữ Việt Nam. Ngoài ra, các phương pháp truyền thống thiếu sự hỗ trợ của ngôn ngữ học khối liệu.

1.2. Cơ sở lý thuyết

1.2.1. Một số khái niệm cốt lõi trong luận án

Trong luận án, “tiếng Hoa” được hiểu là biến thể tiếng Hán được sử dụng trên Chuyên trang *Báo Sài Gòn Giải Phóng* điện tử. “Tiếng Hán phổ thông” được xác định là tiếng Hán chuẩn mực hiện đại dựa trên âm Bắc Kinh, phương ngữ miền Bắc và các tác phẩm văn học bạch thoại kinh điển. “Tiếng Hán cổ” được định nghĩa theo học thuyết phân kỳ của Vương Lực, bao gồm giai đoạn từ thời Tiên Tần đến trước thời Đường (thế kỷ 9). Về phương ngữ, cộng đồng người Hoa tại Việt Nam chủ yếu sử dụng phương ngữ Quảng Đông.

1.2.2. Vấn đề từ vựng

Để đảm bảo tính thống nhất và phù hợp giữa công cụ phân tích với đối tượng nghiên cứu là từ vựng tiếng Hoa, khung lý thuyết của luận án chủ yếu dựa trên từ vựng học tiếng Hán. Các quan điểm từ vựng học tiếng Việt sẽ được tham khảo có chọn lọc nhằm phục vụ cho mục đích đối chiếu trong các phần phân tích sau.

1.2.2.1. Từ, phân loại cấu tạo từ và từ vựng trong tiếng Hán

Khái niệm “từ” trong tiếng Hán được xác định là “đơn vị ngôn ngữ nhỏ nhất có âm, có nghĩa và có khả năng vận dụng độc lập trong ngôn ngữ”. Dựa vào số lượng hình vị và phương thức cấu tạo, “từ” trong tiếng Hán chia thành từ đơn và từ phức. Trong đó, từ đơn gồm từ đơn âm tiết và từ đa âm tiết, từ phức gồm từ ghép, từ láy và từ phái sinh. Từ vựng tiếng Hán gồm từ và ngữ cố định, trong đó, ngữ cố định gồm danh từ riêng, thực ngữ, từ rút gọn và các cụm từ cố định. Luận án sử dụng định nghĩa và hệ thống phân loại trên để phân tích hệ thống từ vựng trên tờ báo.

1.2.2.2. Từ Hán Việt trong tiếng Việt

Từ Hán Việt là một lớp từ vựng quan trọng trong tiếng Việt, hình thành từ quá trình tiếp xúc ngôn ngữ Hán - Việt kéo dài hàng nghìn năm. Theo các nghiên cứu, từ Hán Việt chiếm khoảng 60-80% vốn từ vựng tiếng Việt, đặc biệt trong các lĩnh vực chính trị, khoa học và văn bản hành chính. Về ngữ nghĩa, từ Hán Việt có thể thu hẹp hoặc mở rộng nghĩa so với từ gốc trong tiếng Hán; về cấu trúc, chúng thể hiện sự dịch chuyển cương vị ngữ pháp và thay đổi trật tự hình vị. Trong luận án, từ Hán Việt được dùng làm cơ sở để phân tích sự ảnh hưởng của tiếng Việt đối với tiếng Hoa trên báo SGGP.

1.2.2.3. Hệ thống từ loại trong tiếng Việt và tiếng Hán

Đối chiếu hệ thống từ loại cho thấy tiếng Hán (12 từ loại) và tiếng Việt (8 từ loại) có sự khác biệt đáng kể, đặc biệt trong phạm vi danh từ, kết từ và trợ từ. Luận án sử dụng hệ thống 12 từ loại tiếng Hán làm khung phân tích chính, kết hợp bảng quy chiếu để chuyển đổi các nhãn NLP nhằm đảm bảo tính khách quan và toàn diện trong phân tích ngữ liệu.

1.2.2.4. Nghĩa của từ và cấu trúc nghĩa của từ

Luận án tiếp cận nghĩa của từ như một thực thể tinh thần phức hợp, sử dụng khung phân tích bốn thành phần: nghĩa từ vựng, nghĩa ngữ pháp, nghĩa sắc thái và nghĩa văn hóa, thông qua phương pháp cấu trúc luận với hai đơn vị cơ bản là nghĩa vị và nét nghĩa. Nét nghĩa cấu thành nghĩa vị, nghĩa vị cấu thành nghĩa của từ.

1.2.2. Tiếp xúc ngôn ngữ và biến thể ngôn ngữ

Ngôn ngữ luôn biến đổi dưới ảnh hưởng của các yếu tố xã hội, văn hóa và lịch sử. Ngôn ngữ học xã hội nghiên cứu mối tương quan này, trong đó lý thuyết tiếp xúc ngôn ngữ và biến thể ngôn ngữ đóng vai trò quan trọng trong việc lý giải sự thay đổi của từ vựng, cả từ những tác động bên ngoài lẫn những điều chỉnh bên trong của hệ thống.

1.2.2.1. Tiếp xúc ngôn ngữ

Lý thuyết tiếp xúc ngôn ngữ của Weinreich (1953) khẳng định khi các ngôn ngữ cùng tồn tại trong một cộng đồng, chúng sẽ tương tác và ảnh hưởng lẫn nhau. Trong đó, từ vựng là thành tố đầu tiên chịu tác động thông qua các hiện tượng như vay mượn từ ngữ, mở rộng nghĩa và thâm thấu ngữ pháp. Trên cơ sở này, luận án vận dụng lý thuyết để khảo sát cụ thể các hiện tượng vay mượn và biến đổi từ vựng tiếng Hoa được sử dụng trên Chuyên trang Báo Sài Gòn Giải Phóng điện tử.

1.2.2.2. Biến thể ngôn ngữ

Áp dụng lý thuyết biến thể ngôn ngữ của Hàn Hiếu Minh, tiếng Hoa trên Chuyên trang Báo Sài Gòn Giải Phóng điện tử được xác định là một biến thể của tiếng Hán dựa trên hai tiêu chí: (1) có khác biệt ổn định về từ vựng và cách diễn đạt so với tiếng Hán phổ thông, như “粉紅簿” (sổ đỏ), “广面” (mỳ Quảng); (2) có chức năng xã hội ổn định khi được cộng đồng người Hoa tại Việt Nam sử dụng lâu dài trong giao tiếp và truyền thông.

1.2.3. Ngôn ngữ học đối chiếu

Ngôn ngữ học đối chiếu là một chuyên ngành quan trọng trong ngôn ngữ học, chuyên phân tích các điểm tương đồng và dị biệt giữa các ngôn ngữ để khám phá quy luật ngôn ngữ. Lĩnh vực này không chỉ có giá trị lý luận sâu sắc mà còn có tính ứng dụng cao, đặc biệt trong

các lĩnh vực giảng dạy ngôn ngữ, dịch thuật và nghiên cứu tiếp xúc ngôn ngữ trong môi trường đa ngữ.

1.2.3.1. Các góc độ tiếp cận trong nghiên cứu đối chiếu ngôn ngữ

Luận án sử dụng phương pháp đối chiếu ngôn ngữ trên hai bình diện: đối chiếu nội ngữ (so sánh tiếng Hoa trên báo SGGP với tiếng Hán phổ thông) và đối chiếu liên ngữ (so sánh với từ Hán Việt trong tiếng Việt).

1.2.3.2. Nguyên tắc và phương pháp đối chiếu

Luận án vận dụng phương pháp đối chiếu ngôn ngữ theo quan điểm của Bùi Mạnh Hùng qua ba bước: miêu tả, đối chiếu với tiếng Hán phổ thông, đối chiếu với từ Hán Việt.

1.2.3.3. Nghiên cứu đối chiếu từ vựng

Theo phương pháp đối chiếu từ vựng của Hứa Dư Long, luận án tập trung vào ba trọng tâm ngữ nghĩa: phương thức cấu tạo từ, tính tạo nguyên do và kết hợp từ. Cách tiếp cận này cho phép làm rõ đặc điểm từ vựng tiếng Hoa trong bối cảnh tiếp xúc ngôn ngữ tại Việt Nam.

1.2.4. Ngôn ngữ học khối liệu và nghiên cứu từ vựng dựa trên ngữ liệu

Ngôn ngữ học khối liệu cung cấp phương pháp nghiên cứu hiện đại dựa trên các bộ dữ liệu văn bản số lớn, cho phép phân tích từ vựng một cách khách quan và có hệ thống. Phần này trình bày nền tảng lý thuyết, phương pháp phân tích từ vựng phổ biến, các công cụ chuyên dụng cùng những khái niệm và chỉ số cốt lõi trong xử lý ngữ liệu, đồng thời giới thiệu khối liệu cụ thể được sử dụng trong luận án.

1.2.4.1. Một vài nét về ngôn ngữ học khối liệu

Khối liệu được phân loại theo nhiều tiêu chí như: phạm vi thời gian (đồng đại/ lịch đại), hình thức xử lý (thô/ chủ thích), cấu trúc (cân đối/ ngẫu nhiên), mục đích (tổng quát/ chuyên biệt) và loại hình ngữ liệu (văn bản/ lời nói). Trong luận án này, khối liệu báo SGGP thuộc loại đồng đại - đơn ngữ - chuyên biệt - văn bản viết, đã được xử lý sơ bộ để phục vụ mục tiêu phân tích từ vựng tiếng Hoa trên báo chí.

1.2.4.2. Các khối liệu được sử dụng trong luận án

Luận án sử dụng năm khối liệu gồm: BCC (phần khối liệu báo chí), CCL, LIVAC, Vietlex và khối liệu tự xây dựng từ Chuyên trang *Báo Sài Gòn Giải Phóng* điện tử (khối liệu SGGP Hoa văn), tạo cơ sở đối chiếu toàn diện cho nghiên cứu từ vựng..

1.2.4.3. Các phương pháp chính trong nghiên cứu từ vựng bằng ngôn ngữ học khối liệu

Ngôn ngữ học khối liệu cung cấp bốn phương pháp nghiên cứu từ vựng: phân tích tần suất, kết hợp từ, sắc thái ngữ nghĩa và mối quan hệ ngữ pháp.

1.2.4.4. Các công cụ phân tích dữ liệu dựa trên phương pháp ngôn ngữ học khối liệu

Dựa trên phương pháp ngôn ngữ học khối liệu, luận án sử dụng ba công cụ phân tích: Python với Jieba để xử lý và thống kê từ vựng, AntConc để phân tích kết hợp từ, và LancsBox để trực quan hóa mạng lưới kết hợp.

1.2.4.5. Các khái niệm và chỉ số chính trong ngôn ngữ học khối liệu

Luận án sử dụng các chỉ số: tần suất từ (Freq), điểm T (T-score), điểm tương hỗ thông tin (MI), điểm chuẩn hóa (Z-score), khoảng cách (span) và vị trí (position), với từ trực (node word) làm trung tâm phân tích, áp dụng ngưỡng $\text{Freq} \geq 3$, $\text{MI} \geq 3$, $\text{T-score} \geq 2.33$, $\text{Z} \geq 3$ để xác định kết hợp nổi bật.

Tiểu kết chương 1

Chương 1 của luận án đã hệ thống hóa các nghiên cứu liên quan đến báo *Sài Gòn Giải Phóng* Hoa văn từ góc độ truyền thông báo chí và đối chiếu từ vựng Hán - Hoa, Hán - Việt, qua đó chỉ ra những khoảng trống nghiên cứu mà luận án sẽ giải quyết. Đồng thời chương đã xác lập hệ thống cơ sở lý thuyết gồm từ vựng học, ngôn ngữ học khối liệu, ngôn ngữ học đối chiếu và ngôn ngữ học xã hội, đặt nền móng vững chắc cho các phân tích chuyên sâu ở các chương tiếp theo.

CHƯƠNG 2. DIỆN MẠO TỪ VỰNG TRÊN CHUYÊN TRANG BÁO SÀI GÒN GIẢI PHÓNG ĐIỆN TỬ

2.1. Xây dựng khối liệu từ vựng của Chuyên trang Báo Sài Gòn Giải Phóng điện tử

2.1.1. Phạm vi ngữ liệu

Nguồn ngữ liệu luận án gồm bài viết tự biên từ 9 chuyên mục chính của báo SGGP (3/2017-3/2025), loại bỏ các chuyên mục có tỷ lệ chuyển tải cao từ báo khác và chỉ giữ lại bài viết có ký hiệu “本報訊”, “本報消息” để đảm bảo tính thống nhất về thể loại và phong cách ngôn ngữ.

2.1.2. Quy trình thu thập và xử lý ngữ liệu

2.1.2.1. Quy trình thu thập ngữ liệu

Dữ liệu được thu thập qua 4 bước: trích xuất liên kết, phân tích HTML, phân loại bài viết, lưu thành file .txt.

2.1.2.2. Xử lý ngữ liệu

Việc xử lý ngữ liệu trong luận án được thực hiện qua hai bước chính: tách từ và gán nhãn từ loại bằng thư viện Jieba. Do tính đặc thù của ngữ, quá trình xử lý kết hợp giữa tự động và rà soát thủ công để đảm bảo độ chính xác.

2.1.3. Kết quả thu thập ngữ liệu

Tổng số ngữ liệu nghiên cứu đạt 9.475.036 lượt từ, được thu thập từ 9 chuyên mục: Thời sự chính trị (1.454.093 lượt từ), Pháp luật (937.981 lượt từ), Kinh tế (1.374.960 lượt từ), Nhịp sống người Hoa (941.211 lượt từ), Giáo dục (1.211.823 lượt từ), Thể thao (554.492 lượt từ), Văn hóa giải trí (1.381.942 lượt từ), Độc giả – Từ thiện (1.167.194 lượt từ) và Du lịch (451.340 lượt từ).

2.2. Bức tranh từ vựng trên Chuyên trang Báo Sài Gòn Giải Phóng điện tử

2.2.1. Sự phân bố từ vựng

Luận án đã tiến hành tổng hợp, xử lý dữ liệu đối với toàn bộ ngữ liệu của chín chuyên mục trên Chuyên trang *Báo Sài Gòn Giải Phóng* điện tử. Kết quả thu được là 9.475.036 lượt từ, trong đó có 132.996 đơn vị từ vựng không trùng lặp, gồm từ và ngữ cố định.

2.2.2. Thống kê độ dài âm tiết

Phân tích cho thấy từ vựng được sử dụng trên tờ báo có xu hướng song tiết hóa rõ rệt. Độ dài âm tiết có thể được xem là chỉ số đánh giá mức độ nguyên bản của từng chuyên mục.

2.2.3. Thống kê tần suất từ vựng

Phân tích tần suất cho thấy sự phân bố theo quy luật Zipf. Đối chiếu với bảng tần suất tiếng Hán hiện đại cho thấy nhiều từ thuộc vùng tần suất thấp không có trong từ vựng tiếng Hán hiện đại.

2.3. Diện mạo từ trên Chuyên trang Báo Sài Gòn Giải Phóng điện tử

2.3.1. Từ đơn

2.3.1.1. Đặc điểm chung của từ đơn trong khối liệu báo SGGP

Từ đơn thể hiện đặc điểm nổi bật là sự chi phối của nhóm hư từ về mặt tần suất. Cụ thể, các hư từ như “的” (của), “在” (tại), “是” (là), “了” (rồi)... chiếm tỷ lệ áp đảo trong nhóm từ đơn có tần số cao nhất, với chức năng chính là cấu tạo cú pháp và biểu đạt quan hệ ngữ pháp, hơn là mang nội dung ngữ nghĩa cụ thể. Điểm đáng chú ý là đại từ nhân xưng “我” (tôi) xuất hiện với tần suất cao, phản ánh dấu ấn cá nhân trong lối viết của tờ báo.

2.3.1.2. Phân tích một số trường hợp cụ thể

Phần này tập trung phân tích một số từ đơn điển hình như “郡” (quận), “时” (thì), “获” (hoạch), “逾” (vượt), “须” (cần). Trong đó, “时” được khảo sát theo chiều lịch đại để làm rõ sự biến đổi cách dùng; “获” được phân tích qua kết hợp từ nhằm nhận diện sắc thái nghĩa; còn “须” được tiếp cận theo trường nghĩa để xác định phạm vi nghĩa.

2.3.2. Từ phức

2.3.2.1. Từ ghép

Kết quả cho thấy từ ghép từ tiếng Hoa trên báo SGGP không có sự khác biệt đáng kể so với tiếng Hán phổ thông.

2.3.2.2. Từ láy

Báo SGGP ghi nhận 763 trường hợp từ láy với các dạng chính: AA, AAB, AABB, ABAB và ABB, phản ánh sự đa dạng trong cách thức tạo từ. Dạng AA chiếm ưu thế nhờ tính linh hoạt, trong khi các dạng phức tạp hơn không phổ biến và mang tính cố định hơn. Từ láy

tập trung ở các chuyên mục nguyên bản như Văn nghệ, Nhịp sống người Hoa, Độc giả – Từ thiện, đóng vai trò tu từ quan trọng trong biểu đạt cảm xúc và tái hiện hình ảnh

2.3.2.3. Từ phái sinh

Trên báo SGGP, đa số yếu tố phái sinh giống với tiếng Hán phổ thông, song các phụ tố “者”(người), “化”(hóa) và “症”(chứng) thể hiện đặc điểm riêng: “者” và “化” đóng vai trò cận phụ tố với chức năng cấu tạo từ ổn định; “症” tuy chưa mất nghĩa gốc nhưng có xu hướng trở thành cận hậu tố nhờ việc vị trí cố định và nghĩa khái quát hóa.

2.4. Diện mạo “Ngữ cố định” trên Chuyên trang báo *Sài Gòn Giải Phóng* điện tử

2.4.1. Danh từ riêng

Kết quả khảo sát cho thấy, danh từ riêng trong khối liệu chủ yếu là tên người và tên địa danh. Về tên người, cấu trúc ba âm tiết (ví dụ: “阮文 X”) chiếm ưu thế, phản ánh quy tắc đặt tên truyền thống của người Việt và tuân thủ chuẩn mực văn hóa – xã hội trong văn phong báo chí. Về tên địa danh, tờ báo thể hiện quy ước dịch thuật riêng biệt, khác với tiếng Hán phổ thông, chẳng hạn như “Dubai” được dịch là “杜拜” thay vì “迪拜”, hay “Đắc Lắc” có hai biến thể “德乐” và “达乐” so với cách dịch thống nhất trong tiếng Hán phổ thông là “多乐”, góp phần định hình bản sắc ngôn ngữ độc đáo của tờ báo.

2.4.2. Thục ngữ

Khảo sát thục ngữ trong khối liệu cho thấy sự tồn tại song hành giữa tính chuẩn mực và biến thể địa phương. Về thành ngữ, bên cạnh các đơn vị có cấu trúc ổn định tương đồng với tiếng Hán phổ thông, xuất hiện nhiều biến thể linh hoạt trong nhóm tần suất thấp thông qua việc thay thế thành tố đồng nghĩa (như “奇难杂症” thay cho “疑难杂症”) hoặc đảo trật tự thành tố (như “露宿风餐” thay cho “风餐露宿”). Các quán ngữ và tục ngữ được sử dụng rộng rãi không có khác biệt đáng kể so với tiếng Hán phổ thông. Sự hiện diện của cả ba nhóm thục ngữ này đã góp phần tạo nên tính sinh động và gần gũi trong ngôn ngữ

báo chí, đồng thời phản ánh quá trình địa phương hóa ngôn ngữ trong cộng đồng người Hoa tại Việt Nam.

2.4.3. Đơn vị rút gọn

Hiện tượng rút gọn từ vựng trên báo SGGP phản ánh nguyên tắc tiết kiệm ngôn ngữ trong tiếng Hán, với 294 trường hợp được ghi nhận, chủ yếu là danh từ và động từ. Trong đó xuất hiện một số dạng rút gọn không tồn tại hoặc ít phổ biến trong tiếng Hán phổ thông, như “合绘” (hợp hội/hợp tác hội họa) và “社经” (xã kinh/xã hội kinh tế). Nhiều động từ rút gọn có nguồn gốc từ phương ngữ, ví dụ: “拨电”(gọi điện), “忆述”(nhớ và kể lại), một số khác bắt nguồn từ Hán cổ hoặc chịu ảnh hưởng trực tiếp từ tiếng Việt như “信学”(tin học).

2.4.4. Các cụm từ cố định

Phân tích cho thấy các cụm từ cố định, ví dụ: “总的来说”(nói chung), “越来越”(càng...càng) trên tờ báo hoàn toàn tương đồng với tiếng Hán phổ thông về cấu trúc, ngữ nghĩa và cách sử dụng, không xuất hiện biến thể địa phương. Hiện tượng này phản ánh tính ổn định và phổ quát cao của nhóm từ vựng này trong toàn bộ hệ thống tiếng Hán hiện đại.

2.5. Những đơn vị từ vựng không xuất hiện trong tiếng Hán phổ thông

Trong khối liệu báo SGGP, có một hiện tượng nổi bật là sự xuất hiện của một số đơn vị từ vựng khó tìm thấy trong tiếng Hán hiện đại, và thậm chí không thể đoán được nghĩa nếu người đọc không có kiến thức tiếng Việt hoặc kiến thức về phương ngữ Trung Quốc.

2.5.1. Một số ví dụ tiêu biểu

Khảo sát đã xác định 332 đơn vị từ vựng không tồn tại trong tiếng Hán phổ thông. Phân tích cho thấy các từ vựng này phân bố đa dạng: khoảng 1/4 có yếu tố tương ứng trong tiếng Việt; hơn một nửa có nguồn gốc từ phương ngữ Trung Quốc. Về ngữ nghĩa, một số từ như “蓝勾勾”(dấu tick xanh) có nghĩa không rõ ràng và phụ thuộc vào ngữ cảnh, trong khi các từ như “发梦”(nằm mơ), “勤力”(cần lực) có thể suy đoán nghĩa qua thành phần cấu tạo.

2.5.2. Đặc điểm cấu tạo và ngữ nghĩa

Dựa trên sự miêu tả ở phần trên, luận án đã nhận diện 30 từ vựng nổi bật và tiến hành phân loại chúng thành bốn nhóm chính:

Nhóm A – từ vay mượn từ tiếng Việt; Nhóm B – từ tự tạo; Nhóm C – từ vay mượn từ phương ngữ, nhất là tiếng Quảng Đông; Nhóm D – từ mang yếu tố Hán cổ.

Sự đa dạng này phản ánh quá trình tiếp xúc ngôn ngữ phức hợp, kết hợp giữa bảo tồn truyền thống và thích ứng với môi trường song ngữ Việt – Hán.

Tiểu kết chương 2

Chương 2 đã khảo sát toàn diện hệ thống từ vựng trên báo Sài Gòn Giải Phóng điện tử thông qua hai bước chính: (1) xây dựng và xử lý ngữ liệu chuyên biệt, (2) phân tích định lượng và phân loại từ vựng theo khung toàn diện. Kết quả cho thấy hệ thống từ vựng của báo chủ yếu thể hiện sự “tương đồng lớn hơn dị biệt” so với tiếng Hán phổ thông. Tuy nhiên, điểm dị biệt quan trọng là sự xuất hiện của 332 đơn vị từ vựng không tồn tại trong tiếng Hán phổ thông, phản ánh quá trình tiếp xúc ngôn ngữ và sự thích nghi của tiếng Hán tại Việt Nam. Những phát hiện này tạo cơ sở vững chắc cho các phân tích sâu hơn ở chương tiếp theo.

CHƯƠNG 3. SO SÁNH TỪ VỰNG TRÊN CHUYÊN TRANG BÁO SÀI GÒN GIẢI PHÓNG ĐIỆN TỬ VỚI TIẾNG HÁN PHỔ THÔNG

3.1. Cơ sở phân loại từ vựng

Phân loại từ vựng tiếng Hoa trên báo SGGP được dựa trên các nghiên cứu từ vựng tiếng Hoa tại Đông Nam Á và khảo sát thực tế trong khối liệu. Trên cơ sở đó, luận án đề xuất phân loại từ vựng trên tờ báo thành ba nhóm chính: từ ngữ đặc hữu, từ ngữ dị hình đồng nghĩa và từ ngữ đồng hình dị nghĩa.

3.2. Từ ngữ đặc hữu

3.2.1. Khái niệm và tiêu chí nhận diện

Nhiều học giả đã tiếp cận khái niệm “từ ngữ đặc hữu” theo các hướng khác nhau, nhấn mạnh tính khu vực, xã hội và bản địa hóa trong cộng đồng tiếng Hoa ngoài Trung Quốc. Trong khối liệu, từ ngữ đặc

hữu chủ yếu dùng để biểu đạt các khái niệm, sự vật đặc trưng của xã hội Việt Nam, không tồn tại trong tiếng Hán phổ thông, gồm những đơn vị có cách dịch tương đương nhưng vẫn khác về bản chất, như “祖国阵线” (mặt trận tổ quốc). Luận án xác định từ ngữ đặc hữu là các đơn vị từ vựng mang bản sắc Việt Nam, có tính đặc hữu tương đối mà không xuất hiện trong hệ thống từ vựng tiếng Hán phổ thông.

3.2.2. Phân loại theo chủ đề

Nhằm làm rõ đặc điểm ngữ nghĩa của từ ngữ đặc hữu trên tờ báo, luận án tiến hành phân loại các đơn vị từ vựng đặc hữu đã thu thập theo các chủ đề, cụ thể là: (1) thể chế chính trị - thể chế; (2) ẩm thực; (3) văn hóa – phong tục; (4) xã hội – kinh tế và (5) danh xưng.

3.2.3. Một số ví dụ tiêu biểu

Phần này trình bày một số ví dụ tiêu biểu minh họa cho các chủ đề đã phân loại, chẳng hạn như “白小” (bạc xiu), “企名” (đứng tên) nhằm cụ thể hóa đặc điểm ngữ nghĩa của từ ngữ đặc hữu trên báo SGGP.

3.2.4. Đặc điểm ngữ nghĩa của từ ngữ đặc hữu

Từ ngữ đặc hữu trên tờ báo không chỉ mang nghĩa từ vựng mà còn chứa đựng lớp nghĩa văn hóa sâu sắc, mang tính “phi tương đương” do gắn với bối cảnh văn hóa Việt Nam.

3.3. Từ ngữ dị hình đồng nghĩa so với tiếng Hán phổ thông

3.3.1. Khái niệm và phân loại

“Từ ngữ dị hình đồng nghĩa” được định nghĩa là những đơn vị có cấu trúc hình vị khác nhau (như đảo trật tự, thay thế hình vị) nhưng cùng biểu đạt một khái niệm so với tiếng Hán phổ thông. Dựa trên đặc điểm hình thức, nhóm từ này được chia thành ba loại: (1) khác biệt về lựa chọn hình vị cấu tạo từ; (2) khác biệt về số âm tiết; (3) khác biệt về trật tự hình vị.

3.3.2. Sự khác biệt về việc lựa chọn hình vị cấu tạo từ

Dựa trên mức độ tương đồng hình vị, từ ngữ dị hình đồng nghĩa trên báo SGGP được chia thành hai nhóm: (1) 43 đơn vị có hình vị hoàn toàn khác, ví dụ: “极之”(cực chi) thay “非常”(phi thường), phản ánh ảnh hưởng phương ngữ và thói quen sử dụng đặc thù; (2) 264 đơn

vị có hình vị giống nhau một phần, ví dụ: “借镜”(tá kính) thay “借鉴”(tá giám). Trong đó đa số tuân theo quy tắc “giữ yếu tố hạt nhân, thay yếu tố phụ”. Hiện tượng này cho thấy cơ chế tái tổ hợp từ vựng trong tiếp xúc ngôn ngữ: vừa bảo lưu thành tố ngữ nghĩa chính, vừa thay thế thành tố phụ bằng các yếu tố địa phương hoặc chịu ảnh hưởng từ tiếng Việt, nhằm thích ứng với thói quen biểu đạt của cộng đồng người Hoa tại Việt Nam.

3.3.3. Sự khác biệt về số âm tiết

Sự khác biệt về số âm tiết thể hiện rõ qua hai xu hướng biến đổi âm tiết chính: rút gọn (50 đơn vị) và mở rộng (42 đơn vị). Xu hướng rút gọn, ví dụ: “拨电”(bát điện) thay “打电话”(đá điện thoại/gọi điện) đáp ứng nhu cầu súc tích của báo chí và thể hiện nguyên tắc tiết kiệm trong giao tiếp. Ngược lại, xu hướng mở rộng, ví dụ: “粥水”(chúc thủy) so với “粥”(chúc/cháo) thường nhằm cụ thể hóa ý nghĩa hoặc tăng tính biểu cảm qua cấu trúc lặp. Hiện tượng này không chỉ phản ánh sự linh hoạt trong tiếp biến ngôn ngữ của cộng đồng người Hoa tại Việt Nam mà còn cho thấy sự tác động của khẩu ngữ lên ngôn ngữ viết.

3.3.4. Sự khác biệt về trật tự hình vị

Hiện tượng đảo trật tự hình vị thể hiện qua sự khác biệt về trật tự sắp xếp hình vị so với tiếng Hán phổ thông, chẳng hạn: “bệnh phát”(病发) - “phát bệnh”(发病), “thêm tăng”(添增) - “tăng thêm”(增添). Mặc dù sử dụng chung các hình vị và bảo lưu nghĩa gốc, việc đảo trật tự này đã tạo ra các biến thể diễn đạt độc đáo, phản ánh rõ ảnh hưởng của phương ngữ (đặc biệt là tiếng Quảng Đông), dấu ấn của tiếng Hán cổ.

3.3.5. Đặc điểm ngữ nghĩa của từ ngữ dị hình đồng nghĩa

Từ ngữ dị hình đồng nghĩa trên báo SGGP có những đặc điểm ngữ nghĩa nổi bật. Thứ nhất, các hình vị được thay thế trong quá trình tạo từ thường là những hình vị đồng nghĩa hoặc gần nghĩa, giúp đảm bảo nghĩa gốc của từ không bị thay đổi. Thứ hai, nhiều từ ngữ vừa phản ánh sự bản địa hóa để phù hợp với ngữ cảnh xã hội Việt Nam, vừa bảo lưu yếu tố phương ngữ hoặc từ cổ, thể hiện đặc trưng văn hoá và lịch sử của cộng đồng người Hoa tại Việt Nam. Thứ ba, nhóm từ

này nhìn chung có tính minh bạch ngữ nghĩa cao, nhờ việc giữ lại các yếu tố hình vị chính hoặc bổ sung thêm các thành tố phụ giúp làm rõ nghĩa, tạo thuận lợi cho việc tiếp nhận và sử dụng trong giao tiếp.

3.4. Từ ngữ đồng hình dị nghĩa so với tiếng Hán phổ thông

3.4.1. Khái niệm và phân loại

Từ ngữ đồng hình dị nghĩa là những từ có hình thái cấu trúc giống nhau nhưng khác biệt về nghĩa hoặc cách sử dụng giữa tiếng Hoa trên báo SGGP và tiếng Hán phổ thông. Hiện tượng này được chia thành hai nhóm chính: từ đồng hình dị nghĩa hoàn toàn và từ đồng hình dị nghĩa một phần.

3.4.2. Đồng hình dị nghĩa hoàn toàn

Từ đồng hình hoàn toàn dị nghĩa khá hiếm, chỉ xuất hiện ở một số ít trường hợp, khi cùng một hình thức nhưng mang nghĩa hoàn toàn khác nhau. Trong ngữ liệu, chúng tôi ghi nhận các từ tiêu biểu như “骑士” (kỵ sĩ), “吹管” (ống thổi), “唐山” (Đường Sơn). Đây là những ví dụ điển hình thể hiện sự tách biệt hoàn toàn về mặt ngữ nghĩa trong điều kiện đồng hình.

3.4.3. Đồng hình dị nghĩa một phần

Đây là tiểu loại chiếm tỷ lệ lớn trong nhóm từ đồng hình dị nghĩa. Các đơn vị thuộc nhóm này tuy vẫn giữ một phần nội dung nghĩa gốc trong tiếng Hán, nhưng đã phát sinh biến đổi về nghĩa hoặc cách dùng. Phân tích tập trung vào hai phương diện chính: (1) Biến đổi nghĩa của từ, bao gồm thêm nghĩa mới, giảm nghĩa, mở rộng nghĩa, thu hẹp nghĩa và thay đổi sắc thái nghĩa. (2) Khác biệt về kết hợp từ, thể hiện ở việc thay đổi phạm vi hoặc thói quen kết hợp so với tiếng Hán phổ thông. Phân nghiên cứu sử dụng phương pháp thực chứng với các ví dụ tiêu biểu để minh họa toàn diện cho các hiện tượng biến đổi được nêu ra.

3.4.2. Đặc điểm ngữ nghĩa của từ đồng hình dị nghĩa

Từ đồng hình dị nghĩa trên báo SGGP phản ánh quá trình biến đổi ngữ nghĩa dưới ảnh hưởng của tiếp xúc ngôn ngữ và bối cảnh văn hóa-xã hội Việt Nam. Nhóm từ này có ba đặc điểm chính: (1) Bảo lưu nghĩa gốc: phần lớn từ vẫn giữ lại nghĩa hạt nhân từ tiếng Hán phổ thông; (2) Đặc trưng về ngữ nghĩa thể hiện qua kết hợp từ và (3) Mức độ phụ thuộc vào ngữ cảnh sử dụng cao. Những đặc điểm này cho thấy từ đồng

hình dị nghĩa không chỉ là hiện tượng từ vựng đơn thuần mà phản ánh quá trình tái cấu trúc ngữ nghĩa-ngữ pháp trong môi trường tiếp xúc ngôn ngữ, cung cấp cơ sở quan trọng cho nghiên cứu biến thể tiếng Hán và giảng dạy tiếng Hán trong bối cảnh đa văn hóa.

3.5. Phân tích các yếu tố ảnh hưởng

3.5.1. Yếu tố xã hội

Các yếu tố xã hội tác động mạnh mẽ đến sự biến đổi từ vựng tiếng Hoa, bao gồm: sự kế thừa ngôn ngữ thể hiện ở việc duy trì các yếu tố phương ngữ; Và sự ảnh hưởng của truyền thông và tiếng Hoa trong các khu vực khác, như tuyến thông Hồng Kông, Đài Loan v.v...

3.5.2. Ảnh hưởng từ các ngôn ngữ tiếp xúc

Bên cạnh yếu tố xã hội, tiếp xúc ngôn ngữ cũng đóng vai trò quan trọng, thể hiện qua: ảnh hưởng của phương ngữ Hán và tiếng Hán cổ, giúp bảo tồn nhiều cách dùng từ và cấu trúc cũ; sự ảnh hưởng của tiếng Hán cận đại; và đặc biệt là ảnh hưởng mạnh mẽ từ tiếng Việt, yếu tố tạo ra nhiều thay đổi về nghĩa và cách dùng trong tiếng Hoa trên tờ báo.

3.5.3. Khả năng tạo từ mới của người Hoa Việt Nam

Khối liệu báo SGGP cho thấy khả năng sáng tạo từ vựng của cộng đồng người Hoa tại Việt Nam thông qua các từ đặc hữu như “企名” (xí danh/đứng tên), “转线” (chuyển tuyến), “幼沙” (ấu sa/cát mịn), “白小” (bạc xiu). Những từ này được cấu tạo từ hình vị Hán có sẵn nhưng không tồn tại trong tiếng Hán phổ thông, minh chứng cho “cơ chế sáng tạo nội sinh” nhằm đáp ứng nhu cầu diễn đạt trong xã hội Việt Nam, khẳng định tiếng Hoa tại Việt Nam là hệ thống ngôn ngữ năng động và không ngừng phát triển.

Tiểu kết chương 3

Chương 3 đã phân tích hệ thống từ vựng tiếng Hoa Việt Nam trên SGGP Hoa văn trong sự đối chiếu với tiếng Hán phổ thông qua ba hiện tượng chính: (1) Từ ngữ đặc hữu phản ánh quá trình bản địa hóa và năng lực sáng tạo của cộng đồng người Hoa trong việc định danh các khái niệm đặc thù tại Việt Nam; (2) Từ ngữ dị hình đồng nghĩa minh chứng cho nguồn gốc đa dạng từ phương ngữ Hán, tiếng Hán cổ, tiếng Hán cận đại thông qua các cơ chế cấu trúc như rút gọn, mở rộng âm tiết, thay đổi trật tự hình vị. (3) Từ ngữ đồng hình dị nghĩa thể hiện sự

biến động tinh tế về ngữ nghĩa dưới ảnh hưởng của tiếng Việt và môi trường xã hội; Kết quả cho thấy từ vựng tiếng Hoa Việt Nam là sự kết hợp hài hòa giữa tính kế thừa hệ thống tiếng Hán và tính sáng tạo trong bối cảnh tiếp xúc ngôn ngữ, hình thành nên một biến thể tiếng Hán mang đậm dấu ấn văn hóa Việt Nam.

CHƯƠNG 4. ĐỐI CHIẾU TỪ VỰNG TRÊN CHUYÊN TRANG BÁO SÀI GÒN GIẢI PHÓNG ĐIỆN TỬ VỚI LỚP TỪ HÁN VIỆT

4.1. Khảo sát định lượng và phân tích xu hướng

Phân khảo sát định lượng dựa trên tập dữ liệu 75.894 từ, tiến hành phân tích 3.000 từ thuộc hai vùng tần suất cao và thấp. Kết quả ghi nhận 711 đơn vị từ vựng có từ Hán Việt tương ứng, với tỷ lệ cụ thể là 40.3% ở vùng tần suất cao và 15.4% ở vùng tần suất thấp. Phát hiện này cho thấy hiện tượng từ vựng có từ Hán Việt tương ứng hiện diện một cách phổ biến và có hệ thống trong tiếng Hoa trên tờ báo. Và có xu hướng giảm dần khi tần suất từ giảm.

4.2. Đối chiếu từ vựng

Việc đối chiếu từ vựng được thực hiện trên hai phương diện: trật tự hình vị (phân tích hiện tượng đảo trật tự và nguyên nhân) và nghĩa của từ (bao gồm nghĩa từ vựng và nghĩa ngữ pháp). Theo quan điểm ngôn ngữ học chức năng, luận án xem xét các đặc trưng ngữ pháp như một phần của cấu trúc ngữ nghĩa, đặc biệt trong bối cảnh tiếp xúc ngôn ngữ nơi biến đổi nghĩa thường đi kèm với thay đổi vai trò ngữ pháp.

4.2.1. Đối chiếu trên bình diện trật tự hình vị

4.2.1.1. Kết quả đối chiếu

Trong số 711 từ được khảo sát, có 12 trường hợp đảo trật tự hình vị so với từ Hán Việt, chủ yếu thuộc cấu trúc đẳng lập và chính phụ. Đa số các từ bảo lưu nghĩa gốc sau khi đảo trật tự, trong khi một số ít (như “见证”-“chứng kiến”) có sự thay đổi về nghĩa và chức năng từ loại. Hiện tượng này phản ánh quá trình tiếp biến ngôn ngữ khi các từ vay mượn được điều chỉnh để phù hợp với hệ thống ngữ pháp và tư duy tiếng Việt.

4.2.1.2. Phân tích nguyên nhân gây hiện tượng đảo trật tự hình vị

Hiện tượng đảo trật tự hình vị trong từ Hán Việt bắt nguồn từ ba nguyên nhân chính: yếu tố lịch sử (bảo lưu trật tự cổ như "triển khai"), quy tắc ngữ pháp tiếng Việt: trật tự [trung tâm ngữ+ định ngữ] trong từ ghép chính phụ. Đặc biệt, cấu trúc đẳng lập chiếm đa số tạo điều kiện thuận lợi cho việc hoán đổi vị trí mà vẫn bảo toàn nghĩa của từ.

4.2.2. Đối chiếu trên bình diện nghĩa của từ

Nghĩa từ vựng là ý nghĩa cơ bản và ổn định của từ, phản ánh đặc trưng khách quan của sự vật hiện tượng. Nghiên cứu sử dụng *Từ điển tiếng Hán hiện đại* (bản thứ 7) làm chuẩn tham chiếu, kết hợp ngữ liệu thực tế từ báo SGGP để đối chiếu nghĩa từ vựng giữa tiếng Hoa và từ Hán Việt. Qua đó, các từ được phân thành ba nhóm: nghĩa hoàn toàn giống nhau, nghĩa hoàn toàn khác nhau, và nghĩa giống nhau một phần.

4.2.2.1. Nghĩa từ vựng hoàn toàn giống nhau

Nhóm từ có nghĩa từ vựng hoàn toàn giống nhau chiếm tỷ lệ lớn, thể hiện sự tương đồng về mặt ngữ nghĩa giữa từ vựng tiếng Hoa trên báo SGGP và từ Hán Việt, chẳng hạn như “国家” (quốc gia), “主席” (chủ tịch). Do không có khác biệt đáng kể, nhóm từ này sẽ không được phân tích sâu trong luận án.

4.2.2.2. Nghĩa từ vựng hoàn toàn khác nhau

Trong một số ít trường hợp, từ vựng tiếng Hoa và từ Hán Việt hoàn toàn khác biệt về nghĩa do quá trình cải biến ngữ nghĩa theo thói quen tư duy của người Việt. Ví dụ điển hình là từ “教师” (giáo sư) trong tiếng Hoa chỉ giáo viên nói chung, nhưng từ Hán Việt “giáo sư” lại dùng để chỉ học hàm cao nhất trong hệ thống chức danh khoa học.

4.2.2.3. Nghĩa từ vựng giống nhau một phần

Trong nhóm từ có nghĩa giống nhau một phần, các từ vựng trên báo SGGP và từ Hán Việt tương ứng thể hiện sự giao thoa nghĩa thông qua việc chia sẻ chung một nghĩa vị hoặc sắc thái ý nghĩa., chẳng hạn như “技师” (kỹ sư), “文献” (văn hiến), mà trong phần này chỉ lấy một vài trường hợp tiêu biểu để so sánh phân tích.

4.3. Ảnh hưởng của từ Hán Việt đối với tiếng Hoa Việt Nam qua phân tích trường hợp

Phần này tập trung phân tích cơ chế tác động của từ Hán Việt thông qua trường hợp nhóm từ có nghĩa “tương đồng một phần”. Trên cơ sở danh sách 43 từ vựng đã được xác định, nghiên cứu lựa chọn phân tích sâu một số trường hợp điển hình nhằm làm rõ các hiện tượng ngôn ngữ sinh ra dưới ảnh hưởng của từ Hán Việt như sự thay đổi về kết hợp từ và từ loại.

4.3.1. Sự ảnh hưởng thể hiện qua kết hợp từ

Phần này đã đi sau phân tích 13 trường hợp từ vựng cụ thể để làm rõ cơ chế tác động của từ Hán Việt. Kết quả cho thấy sự ảnh hưởng này thể hiện qua hai hiện tượng chính: (1) mở rộng phạm vi kết hợp do ảnh hưởng từ các tổ hợp tiếng Việt, như “克服后果” (khắc phục hậu quả), (2) xuất hiện các kết hợp đặc thù chỉ tồn tại trong tiếng Hoa Việt Nam như “文化本色” (bản sắc văn hóa).

4.3.2. Sự ảnh hưởng thể hiện qua chuyển đổi từ loại

Phân tích 5 trường hợp điển hình cho thấy từ Hán Việt đã tác động sâu sắc đến hệ thống từ loại của tiếng Hoa Việt Nam. Kết quả cho thấy: Sự chuyển đổi từ loại tuân theo khuôn mẫu ngữ pháp tiếng Việt. Dưới tác động của từ Hán Việt, từ vựng tiếng Hoa Việt Nam trở nên linh hoạt hơn về mặt chức năng. Quá trình này phản ánh sự tiếp nhận và tái cấu trúc đặc điểm ngữ pháp của tiếng Việt vào trong hệ thống tiếng Hoa tại Việt Nam.

4.4. Đặc trưng của từ vựng tiếng Hoa Việt Nam trong mối tương quan với từ Hán Việt

Từ vựng tiếng Hoa Việt Nam hình thành một biến thể độc đáo thông qua ba cơ chế chịu ảnh hưởng của từ Hán Việt: biến đổi ngữ nghĩa, thay đổi kết hợp từ, và chuyển đổi từ loại. Khác với một số biến thể tiếng Hoa trong khu vực, tiếng Hoa Việt Nam phát triển chủ yếu thông qua biến đổi ngữ nghĩa dựa trên hệ thống từ Hán Việt - vốn cùng chia sẻ nguồn gốc tiếng Hán, tạo nên diện mạo ngôn ngữ vừa duy trì tiếng Hán chuẩn mực vừa mang đậm bản sắc Việt Nam.

Tiểu kết chương 4

Chương 4 đã phân tích đối chiếu từ vựng tiếng Hoa trên báo Sài Gòn Giải Phóng và từ Hán Việt, tập trung vào nhóm “nghĩa tương đồng một phần”, đối tượng chịu ảnh hưởng rõ nét nhất từ tiếng Việt. Kết quả

cho thấy ảnh hưởng này thể hiện qua ba hiện tượng chính: biến đổi ngữ nghĩa, khác biệt trong kết hợp từ và chuyển đổi từ loại. Những khác biệt này là kết quả của sự tương tác đa chiều giữa yếu tố nội tại và ngoại tại ngôn ngữ, trong đó ảnh hưởng từ hệ thống từ Hán Việt trong tiếng Việt và sự lựa chọn từ vựng của cộng đồng người Hoa tại Việt Nam đóng vai trò then chốt, góp phần làm sáng tỏ đặc trưng của biến thể tiếng Hoa Việt Nam trong khu vực Đông Nam Á.

KẾT LUẬN

Luận án đã tiến hành khảo sát và phân tích đặc điểm từ vựng trên Chuyên trang báo Sài Gòn Giải Phóng điện tử dựa trên khối liệu gồm hơn 9 triệu lượt từ, được thu thập từ 9 chuyên mục với tổng số 132.997 đơn vị từ vựng. Thông qua việc so sánh với tiếng Hán phổ thông và đối chiếu với lớp từ Hán Việt, luận án đã chỉ ra những đặc điểm nổi bật của hệ thống từ vựng tiếng Hoa trên tờ báo, đồng thời làm rõ các yếu tố chi phối sự biến đổi của từ vựng trong môi trường tiếp xúc ngôn ngữ - văn hóa tại Việt Nam.

Về sự khác biệt giữa tiếng Hoa trên báo SGGP với tiếng Hán phổ thông, luận án đã xác định và phân loại được 918 đơn vị từ vựng có sự khác biệt, trong đó, phân bố cụ thể như sau: (1) Từ ngữ đặc hữu có 332 đơn vị, tập trung chủ yếu vào các lĩnh vực văn hóa, xã hội và ẩm thực đặc thù của Việt Nam như “薄寮公子” (công tử Bạc Liêu), “广面” (Mỳ Quảng); (2) Từ ngữ dị hình đồng nghĩa có 307 đơn vị, chẳng hạn như các cặp từ “利是-红包 (lì xì)”, “借镜-借鉴 (lấy làm gương)”, thể hiện sự khác biệt về hình vị, số lượng âm tiết, trật tự từ; (3) Từ ngữ đồng hình dị nghĩa (cùng hình thức nhưng khác nghĩa) có 275 đơn vị, thể hiện sự khác biệt về số lượng nghĩa vị, phạm vi nghĩa, sắc thái biểu cảm và khả năng kết hợp từ, điển hình như “唐山” (Đường Sơn) trong tiếng Hoa Việt Nam chỉ “nơi đất tổ của người Hoa”, trong khi tiếng Hán phổ thông chỉ tên gọi của một thành phố ở tỉnh Hà Bắc Trung Quốc, hoặc từ “素质” (tố chất) trong tiếng Hán phổ thông chủ yếu chỉ bản chất vốn có của sự vật, năng lực tổng hợp cá nhân, đặc điểm sinh lý thần kinh bẩm sinh, mà trong tiếng Hoa trên tờ báo từ này còn thêm

nghĩa vị là “chất lượng”.

Về ảnh hưởng của lớp từ Hán Việt, luận án chỉ ra rằng nhóm từ thuộc diện “nghĩa từ vụng tương đồng một phần” là đối tượng chịu ảnh hưởng của từ Hán Việt rõ nét nhất, đồng thời cũng là nhóm tiêu biểu nhất cho quá trình tiếp xúc ngôn ngữ. Ví dụ điển hình như từ “văn hiến” trong tiếng Việt có nghĩa là “truyền thống văn hóa lâu đời của dân tộc”, và thậm chí có thể dùng cho cả tư liệu khảo cổ, kiến trúc, mỹ thuật. Còn từ “文献” trong tiếng Hán phổ thông chủ yếu dùng để chỉ các tài liệu thành văn, chẳng hạn như “参考文献” (văn hiến tham khảo, tức là tài liệu tham khảo). Mà trong tiếng Hoa trên báo SGGP, từ “文献” vừa có thể chỉ truyền thống văn hóa lâu đời của dân tộc, và cũng có thể chỉ tài liệu thành văn. Cho thấy hệ quả ảnh hưởng của từ Hán Việt đối với tiếng Hoa trên tờ báo.

Về nguyên nhân dẫn đến sự khác biệt trên, luận án xác định ba nhóm nhân tố chi phối: (1) Yếu tố xã hội, chủ yếu là sự kế thừa ngôn ngữ trong gia đình người Hoa và ảnh hưởng của truyền thông tiếng Hoa trong khu vực. Nhiều gia đình người Hoa tại Việt Nam vẫn duy trì thói quen sử dụng phương ngữ quê hương trong đời sống hàng ngày, đặc biệt là trong giao tiếp gia đình và nội bộ cộng đồng. Sự kế thừa ngôn ngữ này đã tạo điều kiện cho các phương ngữ như Quảng Đông, Phúc Kiến, Triều Châu được bảo tồn qua nhiều thế hệ. Bên cạnh đó, tiếng Hoa trên báo SGGP còn chịu tác động từ các khu vực nói tiếng Hoa khác như Hồng Kông, Đài Loan, Singapore và Malaysia thông qua các phương tiện truyền thông. Điều này lý giải vì sao một số từ ngữ trên báo SGGP cũng xuất hiện trong tiếng Hoa của Hồng Kông, Đài Loan, phản ánh sự giao lưu và ảnh hưởng lẫn nhau trong không gian tiếng Hoa toàn cầu. (2) Sự tiếp xúc ngôn ngữ. Sự tiếp xúc này diễn ra trên ba bình diện. Thứ nhất, tiếp xúc với các phương ngữ Hán, đặc biệt là tiếng Quảng Đông, vốn được bảo lưu mạnh mẽ trong cộng đồng người Hoa tại Việt Nam. Tiếng Quảng Đông đã bảo tồn một số lượng lớn từ ngữ của tiếng Hán cổ, và những từ ngữ này được duy trì trong tiếng Hoa Việt Nam như “生果” (sinh quả), “水喉” (thủy hầu). Thứ hai, tiếp xúc với tiếng Hán cận đại, khi nhiều từ ngữ của tiếng Hán

thời cận đại vẫn được lưu giữ trong tiếng Hoa Việt Nam trong khi đã biến mất hoặc ít dùng trong tiếng Hán phổ thông như từ “拳拳” (quyền quyền/tha thiết). Thứ ba, tiếp xúc với tiếng Việt, dẫn đến hiện tượng vay mượn từ vựng và chuyển nghĩa, như “寄车” (gửi xe), “劳力” (lao lực). Những từ ngữ này là kết quả của quá trình người Hoa thích ứng với môi trường ngôn ngữ bản địa, tiếp nhận và chuyển hóa các khái niệm trong tiếng Việt sang tiếng Hoa.

So với tiếng Hán phổ thông, hệ thống từ vựng tiếng trên báo SGGP thể hiện sự tương thích cao trong các chủ đề như chính trị, kinh tế, xã hội, thể hiện rõ “tương đồng lớn hơn dị biệt”, phản ánh xu thế hội tụ và hợp nhất trong mạng lưới tiếng Hoa toàn cầu.

Tuy nhiên, như Điều Yến Bân (刁晏斌) [87] đã chỉ ra, sức sống dị biệt trong hệ thống tiếng Hoa toàn cầu lại bắt nguồn chính từ những “biến thể bên lề” – nơi hội tụ các cơ chế sáng tạo ngôn ngữ và con đường biểu đạt pha trộn. Chính những ngữ liệu tương chừng “phi chuẩn” này đã cung cấp cho chúng ta một cửa sổ kho báu để nhận diện cách thức ngôn ngữ sáng tạo nên cấu trúc nghĩa và mô hình biểu đạt mới trong các ngữ cảnh sử dụng thực tế. Luận án không chỉ dừng lại ở việc thống kê và phân loại sự khác biệt về nghĩa của từ, mà còn thông qua chính sự khác biệt đó để chỉ ra rằng: trong hệ thống tiếng Hoa toàn cầu, tính đa dạng và sức sáng tạo là hai đặc điểm tồn tại song song. Sự đồng nhất ngôn ngữ không có nghĩa là phải loại bỏ sự khác biệt, mà là quá trình tiếp nhận những khả năng biểu đạt dị biệt trong một khuôn khổ mang tính quy chuẩn tương đối. Quá trình này cho thấy, “tiếng Hoa toàn cầu” không phải là kết quả phát triển ngôn ngữ trong trạng thái tĩnh, mà là một hệ thống ngôn ngữ luôn vận động, luôn điều chỉnh và tiến hóa trong môi trường đa ngôn ngữ.

Về đóng góp của luận án, nghiên cứu đã cung cấp một bức tranh toàn diện về đặc điểm từ vựng của một tờ báo tiếng Hoa tại Việt Nam, góp phần lấp đầy khoảng trống nghiên cứu về biến thể tiếng Hán tại khu vực Đông Nam Á nói chung và Việt Nam nói riêng. Kết quả nghiên cứu có giá trị tham khảo đối với công tác biên soạn từ điển tiếng Hoa toàn cầu cũng như việc bổ sung minh chứng cho sự vận động và biến đổi tiếng Hoa trong môi trường xã hội Việt Nam.

Bên cạnh những kết quả đạt được, luận án vẫn còn một số hạn chế: khối liệu chỉ giới hạn trong phạm vi báo chí viết, chưa bao quát được hệ thống khẩu ngữ của tiếng Hoa Việt Nam; dung lượng và phạm vi khảo sát còn hạn chế nên một số hiện tượng từ vựng mang tính thời điểm hoặc địa phương chưa được phát hiện đầy đủ. Hướng nghiên cứu tiếp theo có thể mở rộng phạm vi khảo sát sang các thể loại văn bản khác, kết hợp điều tra khẩu ngữ để có cái nhìn toàn diện hơn về tiếng Hoa Việt Nam, đồng thời tiếp tục hoàn thiện các thao tác phân tích ngữ nghĩa, góp phần thúc đẩy nghiên cứu tiếng Hoa Việt Nam trở thành một mảng chuyên sâu trong khung lý thuyết tiếp xúc ngôn ngữ và biến thể tiếng Hán toàn cầu.

DANH MỤC CÁC CÔNG TRÌNH KHOA HỌC CỦA TÁC GIẢ LIÊN QUAN ĐẾN LUẬN ÁN

1. 李丽娟 (2024), 越南华文报刊华语与普通话的词汇差异初探, Kỷ yếu hội thảo khoa học quốc tế Ngôn ngữ học đối chiếu và đối chiếu giữa các ngôn ngữ. NXB Thế giới, tr.1587-1592, ISBN: 9786043923773.
2. 李丽娟 (2025), 词语搭配视域下“力量”的语义特征研究, 学知 (06), p.52-55, ISSN:3029-0740.
3. 阮黄英, 李丽娟 (2025), 基于词表的两岸词汇差异比较及对越汉语教学. 新时代两岸华语教学国际学术研讨会论文集. tr. 50-63. ISBN: 9789865265724.
4. 李丽娟 (2025), 语料库在国际中文词汇教学中的应用研究——基于越南华文报刊语料的实证分析. 数字时代汉语教学与研究论文集. tr.495-501. ISBN: 9786326074161.
5. Lý Lê Quyên (2025), Hiện tượng vay mượn, bảo lưu và sáng tạo từ vựng trên Báo Sài Gòn Giải Phóng Hoa văn, Tạp chí Ngôn ngữ (08), tr.22-30, ISSN: 0866-7519.