

NGÔN NGỮ HỌC VÀ VIỆT NGỮ HỌC

HỆ THỐNG NHÂN NGỮ NGHĨA THUẬT NGỮ THÂN TỘC TIẾNG VIỆT

PHAN THỊ MỸ TRANG* - ĐẶNG NGỌC LỆ** - ĐINH ĐIỀN*** -
TRẦN THỊ MINH PHƯỢNG****

TÓM TẮT: Thân tộc là đề tài quen thuộc trong nghiên cứu ngôn ngữ và văn hóa. Tuy nhiên, hiện nay chưa có hệ thống nhân ngữ nghĩa cho thuật ngữ thân tộc tiếng Việt dưới góc độ ngôn ngữ học. Nghiên cứu này sử dụng phương pháp thống kê mô tả và phương pháp phân tích - tổng hợp nhằm mô tả, phân loại và hệ thống các từ ngữ thân tộc tiếng Việt dựa trên cấu trúc phân cấp ý niệm và các mối quan hệ ngữ nghĩa như đồng nghĩa, đa nghĩa, thượng danh và hạ danh. Kết quả nghiên cứu mở ra tiềm năng ứng dụng trong giảng dạy tiếng Việt cho người nước ngoài, xử lý ngữ nghĩa tự động và nghiên cứu ngữ nghĩa học so sánh.

TỪ KHÓA: thuật ngữ thân tộc tiếng Việt; hệ thống nhân ngữ nghĩa; đồng nghĩa; đa nghĩa; cấu trúc phân cấp ý niệm.

NHẬN BÀI: 06/11/2024.

BIÊN TẬP-CHỈNH SỬA-DUYỆT ĐĂNG: 11/12/2024

1. Giới thiệu

Nhân ngữ nghĩa thường là các từ hoặc ngữ được sử dụng để mô tả và phân loại ý nghĩa của từ vựng trong ngôn ngữ. Nhân ngữ nghĩa giúp hệ thống hóa và xác định mối quan hệ ý nghĩa giữa các từ, ngữ, hoặc đơn vị ngôn ngữ khác. Ngôn ngữ học hiện đại nhìn nhận từ vựng-ngữ nghĩa là một hệ thống (Lê Quang Thiêm, 2018), theo đó, việc phân loại và tổ chức theo cấu trúc phân cấp với các phạm trù hạ danh và thượng danh là rất quan trọng để xây dựng hệ thống nhân ngữ nghĩa từ vựng. Trong lĩnh vực ngữ nghĩa học, các hệ thống nhân ngữ nghĩa đã và đang đóng vai trò quan trọng trong việc tổ chức và truy xuất thông tin ngữ nghĩa. Sự phát triển của các hệ thống nhân ngữ nghĩa tạo điều kiện thuận lợi cho việc hỗ trợ tìm kiếm thông tin theo trường ngữ nghĩa. Các nghiên cứu trước đây về hệ thống nhân ngữ nghĩa chủ yếu tập trung vào ngôn ngữ tiếng Anh như: WordNet (Fellbaum, 2005), LLOCE (McArthur, 1988), UCREL (Rayson và các tác giả khác, 2004). Trong tiếng Việt, mặc dù đã có một số nghiên cứu về thuật ngữ thân tộc từ góc độ văn hóa và xã hội, tuy nhiên các nghiên cứu về hệ thống nhân ngữ nghĩa trong tiếng Việt vẫn còn hạn chế. Hệ thống thuật ngữ thân tộc tiếng Việt khá đa dạng và độc đáo, không chỉ mang ý nghĩa về mặt huyết thống, hôn nhân mà còn thể hiện giá trị văn hóa, xã hội sâu sắc. Việc xây dựng hệ thống nhân ngữ nghĩa cho thuật ngữ thân tộc trong tiếng Việt dưới góc độ ngôn ngữ học không chỉ giúp tạo ra một hệ thống hoàn toàn dựa trên ngữ liệu tiếng Việt, làm rõ cấu trúc phân cấp và các mối quan hệ ngữ nghĩa mà còn hỗ trợ cho các lĩnh vực ứng dụng khác như: giảng dạy tiếng Việt cho người nước ngoài, xử lý ngữ nghĩa tự động và nghiên cứu ngữ nghĩa học từ vựng.

Mục tiêu của nghiên cứu này là xây dựng hệ thống nhân ngữ nghĩa cho thuật ngữ thân tộc tiếng Việt, tập trung vào cấu trúc phân cấp và các mối quan hệ ngữ nghĩa chính bao gồm đồng nghĩa, đa nghĩa, thượng danh và hạ danh.

Nghiên cứu này nhằm trả lời các câu hỏi sau:

1. Hệ thống nhân ngữ nghĩa cho thuật ngữ chỉ quan hệ thân tộc trong tiếng Việt được tổ chức như thế nào?

2. Các mối quan hệ đồng nghĩa, đa nghĩa, thượng danh và hạ danh được xác định như thế nào trong hệ thống nhân ngữ nghĩa thuật ngữ thân tộc tiếng Việt?

2. Lịch sử nghiên cứu

WordNet là một hệ thống từ vựng tiếng Anh nổi bật, được phát triển vào năm 1986 bởi nhóm các nhà khoa học dưới sự dẫn dắt của giáo sư tâm lý học George Armitage Miller tại Đại học Princeton. Ban đầu, WordNet có 25 ý niệm gốc nhưng sau đó được rút gọn thành 11 ý niệm nguyên thủy

*Trường Đại học Công nghệ Sài Gòn; Email: mytrang779@gmail.com

**PGS.TS; Trường Đại học Nguyễn Tất Thành; Email: ngocle.dang1309@gmail.com

***PGS.TS; Trường Đại học Khoa học Tự nhiên, ĐHQG TP. Hồ Chí Minh;

Email: ddiem@fit.hcmus.edu.vn

****TS; Trường Đại học Khoa học xã hội và Nhân văn - ĐHQG TP. Hồ Chí Minh;

Email: minhphuongtrn@hcmussh.edu.vn

(primitive concept) bao trùm tất cả danh từ trong tiếng Anh. Các loại từ gồm danh từ, động từ, tính từ và trạng từ được nhóm thành các tập đồng nghĩa (synonym sets; synsets), mỗi tập đại diện cho một ý niệm riêng biệt [Miller và các tác giả khác, 1990]. Các tập từ được kết nối với nhau thông qua quan hệ ngữ nghĩa-ý niệm và từ vựng (conceptual-semantic and lexical relations). Trong WordNet, các mối quan hệ ngữ nghĩa giữa các từ bao gồm quan hệ đồng nghĩa (synonymy), quan hệ trái nghĩa (antonymy), quan hệ thuộc cấp (hyponymy), quan hệ bao hàm (hypernymy), quan hệ bộ phận (meronymy), quan hệ tổng thể (holonymy), quan hệ kéo theo (entailment) và quan hệ cách thức đặc biệt (troponymy) [Fellbaum, 1998]. Điểm nổi bật của WordNet là sự liên kết từ không chỉ ở dạng từ mà còn ở cả nghĩa cụ thể của từ, giúp phân biệt các từ gần nhau trong mạng từ về mặt ngữ nghĩa. WordNet cũng gán nhãn cho các mối quan hệ ngữ nghĩa giữa các từ. Hiện tại, WordNet chứa khoảng 117.000 lớp synsets [Fellbaum, 2005].

Cho đến nay, nhiều quốc gia đã tiến hành xây dựng mạng từ tương tự cho ngôn ngữ của họ. Một trong những dự án nổi bật là EuroWordNet, được khởi động vào năm 1996, nhằm phát triển mạng từ cho các ngôn ngữ trong khối Liên minh châu Âu (EU) như tiếng Hà Lan, Tây Ban Nha và Italia/Ý; đồng thời liên kết các mạng từ này với mạng từ tiếng Anh trong một cơ sở dữ liệu đa ngôn ngữ. Đến năm 1997, dự án đã mở rộng sang tiếng Đức, tiếng Pháp, tiếng Séc, tiếng Estonia và hoàn thành vào cuối năm 1999. Cơ sở dữ liệu EuroWordNet tương thích cốt lõi với Mạng từ tiếng Anh, đã được bổ sung và mở rộng đáng kể về các mối quan hệ và đối tượng (Vossen, 2002). Trong các công trình nghiên cứu tiếng Việt, Trương Hoàng Bảo và các cộng sự (2021) đã xây dựng WordNet tiếng Việt tự động dựa trên ngữ liệu song ngữ. Nghiên cứu đã cải tiến phương pháp chọn lọc các tập đồng nghĩa bằng cách sử dụng kho ngữ liệu TEDTalk và chỉ dựa vào WordNet của bốn ngôn ngữ nguồn: Anh, Tây Ban Nha, Pháp và Nhật Bản. Kết quả cho thấy phương pháp đạt độ chính xác là 61,42%, xây dựng được WordNet tiếng Việt với 16.143 từ và 17.756 synsets. Phương pháp này hoàn toàn có thể được áp dụng để xây dựng WordNet cho các ngôn ngữ khác, đồng thời bổ sung thêm từ vựng cũng như synsets vào WordNet của các ngôn ngữ nguồn. Tiếp theo, Lam và K. N. & Kalita, J. (2022) đã phát triển Mạng từ tiếng Việt (VWN) từ các nguồn miễn phí của nhiều ngôn ngữ khác nhau, với cấu trúc tương tự như Mạng từ tiếng Anh (Princeton WordNet). Nhóm nghiên cứu đã sử dụng thuật toán tự động để dịch các từ tiếng Anh trong WordNet sang tiếng Việt, đồng thời loại bỏ sự nhập nhằng trong kết quả dịch bằng cách xếp hạng các từ dựa trên số lần xuất hiện và khoảng cách tìm kiếm trên Google. Bên cạnh đó, nhóm tác giả cũng liên kết các tập đồng nghĩa (synsets) và trích xuất các định nghĩa (glosses) cho các synsets. Ngoài ra, nhóm tác giả đã xem xét những vấn đề phát sinh trong VWN do sự khác biệt về văn hóa và hình thái đơn lập của tiếng Việt so với các ngôn ngữ khác [Lam, K. N. & Kalita, J., 2022)]. Nhìn chung, các công trình nghiên cứu trên nhấn mạnh tầm quan trọng của tài nguyên từ vựng ngữ nghĩa và việc tổ chức, hệ thống nhãn ngữ nghĩa trong phân tích ngôn ngữ và xử lý ngôn ngữ tự nhiên, mở ra tiềm năng lớn cho ứng dụng thực tế. Bên cạnh đó, các công trình nghiên cứu này cung cấp cơ sở khoa học và làm nổi bật tính khả thi của việc xây dựng hệ thống nhãn ngữ nghĩa và phát triển các công cụ chú thích ngữ nghĩa cho các ngôn ngữ mới.

Trong nghiên cứu về quan hệ thân tộc trong tiếng Việt, Dương Thị Nụ (2003) tập trung vào việc đối chiếu đặc trưng ngữ nghĩa của nhóm từ chỉ quan hệ thân tộc trong tiếng Anh và tiếng Việt. Công trình này không chỉ nhằm tìm hiểu các nghĩa cơ bản và mở rộng của từ thân tộc trong hai ngôn ngữ mà còn hướng tới việc cải thiện hiệu quả giao tiếp liên văn hóa giữa người Việt và người Anh nói tiếng Anh - Anh và tiếng Anh - Mỹ. Đề tài của tác giả tập trung vào ba thế hệ trên Ego (bản thân) (thế hệ cha mẹ G-1, thế hệ ông bà G-2 và thế hệ cụ G-3) và ba thế hệ dưới Ego (bản thân) (thế hệ con G+1, thế hệ cháu G+2 và thế hệ chắt G+3), cùng những người cùng thế hệ với Ego (bản thân). Nghiên cứu này làm sáng tỏ sự tương đồng và khác biệt về đặc trưng ngữ nghĩa của nhóm từ thân tộc, từ đó góp phần vào lý thuyết ngôn ngữ học đại cương về tính tương đối của ngôn ngữ, tính có ý nghĩa của ký hiệu ngôn ngữ và chức năng giao tiếp của ngôn ngữ [Dương Thị Nụ, 2003]. Nguyễn Văn Khang (2014) tập trung vào việc sử dụng các từ xưng hô theo quan hệ thân tộc trong giao tiếp công quyền, với mục đích tìm hiểu liệu việc chuẩn hóa các từ xưng hô này trong giao tiếp công quyền có khả thi

hay không. Nghiên cứu cho thấy rằng các từ xưng hô theo quan hệ thân tộc không chỉ chiếm một phần lớn trong hệ thống từ xưng hô tiếng Việt, mà còn là nội dung cốt lõi trong việc chuẩn hóa giao tiếp công quyền [Nguyễn Văn Khang, 2014]. Nguyễn Võ Tâm Như (2023) tập trung nghiên cứu và phân tích các từ ngữ chỉ quan hệ thân tộc trong hai ngôn ngữ Việt - Nhật dựa trên lý thuyết lược đồ vai trò. Mục đích của nghiên cứu nhằm làm rõ sự tương đồng và dị biệt trong cấu trúc ngữ nghĩa, cách sử dụng từ ngữ chỉ quan hệ thân tộc giữa hai nền văn hóa. Bài viết sử dụng phương pháp thống kê, so sánh và đối chiếu, thông qua việc khảo sát 119 từ chỉ quan hệ thân tộc trong tiếng Việt và 156 từ trong tiếng Nhật. Kết quả nghiên cứu góp phần giải mã những khái niệm văn hóa trong hai cộng đồng, cung cấp ứng dụng thực tiễn trong việc giảng dạy, học tập tiếng Việt và tiếng Nhật, đặc biệt trong giao tiếp liên văn hóa.

3. Phương pháp nghiên cứu

Trong nghiên cứu hệ thống nhãn ngữ nghĩa cho thuật ngữ thân tộc tiếng Việt, chúng tôi tiến hành thu thập ngữ liệu, sử dụng phương pháp thống kê mô tả và phương pháp phân tích - tổng hợp.

3.1. Thu thập ngữ liệu

Ngữ liệu về các thuật ngữ thân tộc trong tiếng Việt được thu thập bao gồm các từ và ngữ được xác lập thông qua mối quan hệ dòng tộc, hôn nhân và gia đình. Các thuật ngữ này được trích dẫn từ nhiều nguồn tài liệu như:

- Các công trình nghiên cứu trước đây về thuật ngữ thân tộc của các tác giả: Dương Thị Nụ (2003), Nguyễn Văn Khang (2014) và Nguyễn Võ Tâm Như (2023).

- Từ điển tiếng Việt của Viện Ngôn ngữ học do Hoàng Phê chủ biên (tái bản năm 2020). Đây là nguồn tài liệu chính thức để liệt kê các thuật ngữ thân tộc trong tiếng Việt.

- Ngữ liệu được trích dẫn từ các nguồn từ điển khác nhằm mở rộng thêm số lượng thuật ngữ, hỗ trợ việc xây dựng các mối quan hệ ngữ nghĩa trong hệ thống nhãn ngữ nghĩa thuật ngữ thân tộc tiếng Việt, bao gồm: Từ điển phương ngữ tiếng Việt của tác giả Phạm Văn Hảo (2009); Từ điển phương ngữ tiếng Việt của tác giả Đặng Thanh Hòa (2005); Từ điển đối chiếu Từ địa phương của các tác giả Nguyễn Như Ý, Đặng Ngọc Lệ và Phan Xuân Thành (2001); Từ điển Từ ngữ Nam Bộ của tác giả Huỳnh Công Tín (2007); Từ điển trái nghĩa - đồng nghĩa tiếng Việt của các tác giả Dương Kỳ Đức và Vũ Quang Hào (1994).

Ngoài ra, chúng tôi còn trích dẫn ngữ liệu từ các tác phẩm văn học (truyện ngắn, tiểu thuyết), báo chí và các tài liệu chính thống khác của: Nhà xuất bản Văn học, Văn học Sài Gòn, Hội Văn học Nghệ thuật tỉnh Quảng Ngãi, Sông Thu Xuất bản, Hội Nhà Văn Việt Nam, Tạp chí Sông Hương, Tạp chí Văn nghệ Thái Nguyên điện tử, Tạp chí Cửa Việt, Nhịp sống Hà Nội - Chuyên trang của Báo điện tử Hà Nội mới, Báo Thanh niên, Báo Tiền Phong, Báo Giác ngộ (online), Báo Giáo dục và thời đại, Báo Điện tử Phụ nữ Thành phố Hồ Chí Minh, Báo Tuổi trẻ online, Trang web của Công ty luật TNHH Dương Gia, Báo điện tử Pháp luật Việt Nam, Báo điện tử Quảng Ninh, Báo điện tử Tiền Phong, Báo điện tử Pháp Luật Thành phố Hồ Chí Minh, Báo điện tử VnExpress, Báo giáo dục và thời đại, Báo Phụ nữ Việt Nam Điện tử và Báo Nhân dân.

3.2. Phân loại ngữ liệu

Các thuật ngữ thân tộc thu thập sẽ được phân loại theo các bậc như: bậc kị, bậc cụ, bậc ông bà, bậc cha mẹ, bậc bác chú cô cậu gì/ dì, bậc vợ chồng, bậc anh chị em, bậc con cháu.

3.3. Xây dựng cấu trúc phân cấp ý niệm

Sau khi hoàn tất việc phân loại ngữ liệu, chúng tôi tiến hành xây dựng hệ thống nhãn ngữ nghĩa cho thuật ngữ thân tộc trong tiếng Việt dựa trên cấu trúc phân cấp theo nghĩa ý niệm (Leech, 1985, tr.23). Cấu trúc này sẽ được tổ chức từ khái niệm tổng quát đến các khái niệm cụ thể, dựa trên cách tổ chức các trường ngữ nghĩa và các mối quan hệ ngữ nghĩa giữa các thuật ngữ.

3.4. Phân tích mối quan hệ ngữ nghĩa

Cuối cùng, chúng tôi sẽ phân tích mối quan hệ ngữ nghĩa trong hệ thống nhãn ngữ nghĩa cho thuật ngữ thân tộc tiếng Việt bao gồm: đồng nghĩa (synonymy), đa nghĩa (polysemy), thượng danh (hyponymy) và hạ danh (hyponymy).

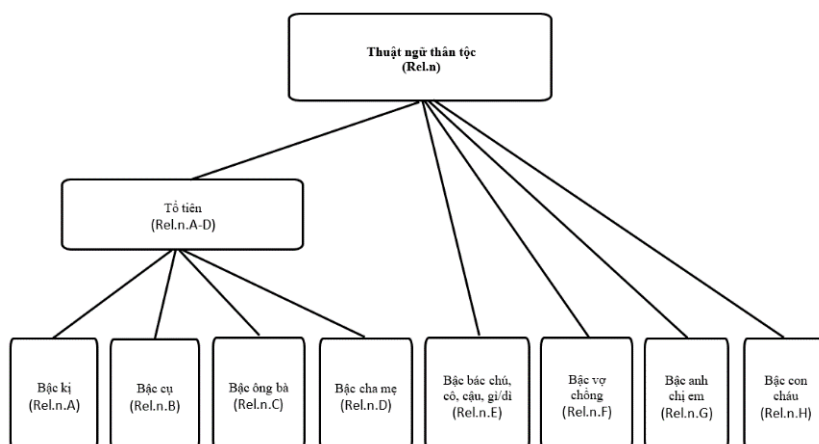
4. Kết quả nghiên cứu và thảo luận

Trong phần này, chúng tôi trình bày những kết quả chính của nghiên cứu về hệ thống nhãn ngữ nghĩa cho các thuật ngữ thân tộc trong tiếng Việt. Bắt đầu với việc mô tả cấu trúc phân cấp của hệ thống, chúng tôi sẽ cung cấp thông tin chi tiết về số lượng thuật ngữ, mã nhãn ngữ nghĩa của cả hệ thống và của riêng mỗi bậc; sau đó phân tích các mối quan hệ ngữ nghĩa quan trọng như đồng nghĩa, đa nghĩa, thượng danh và hạ danh. Ngoài ra, chúng tôi cũng đề xuất một số ứng dụng thực tiễn cho hệ thống nhãn ngữ nghĩa này nhằm làm rõ hơn vai trò của hệ thống đối với việc sử dụng ngôn ngữ về thân tộc trong văn hóa Việt Nam.

4.1. Cấu trúc phân cấp hệ thống nhãn ngữ nghĩa cho thuật ngữ thân tộc tiếng Việt

Hệ thống nhãn ngữ nghĩa cho thuật ngữ thân tộc tiếng Việt của chúng tôi được xây dựng dựa trên cấu trúc phân cấp ý niệm, tổng cộng có 307 thuật ngữ thân tộc và 201 nhãn ngữ nghĩa, bao gồm khái niệm tổng quát, lớp trung gian và các từ/ ngữ thực tế. Hệ thống bắt đầu từ ý niệm gốc “thuật ngữ thân tộc” được mã hóa với nhãn ngữ nghĩa tổng quát là “Rel.n”, trong đó “Rel” là dạng viết tắt của “relative” (họ hàng), “n” là kí hiệu cho các danh từ và danh ngữ.

Nhãn con cho lớp trung gian tiếp theo thể hiện mối quan hệ với tổ tiên, được đặt tên nhãn là “tổ tiên” với mã nhãn là Rel.n.A-D. Các nhánh chính của hệ thống ý niệm bao gồm các bậc như: bậc kị (Rel.n.A), bậc cụ (Rel.n.B), bậc ông bà (Rel.n.C), bậc cha mẹ (Rel.n.D), bậc bác chú, cô, cậu, gi/ di (Rel.n.E), bậc vợ chồng (Rel.n.F), bậc anh chị em (Rel.n.G), bậc con cháu (Rel.n.H) (xem Hình 1).



Hình 1. Minh họa cấu trúc phân cấp các lớp ý niệm cho “thuật ngữ thân tộc” tiếng Việt

Trong đó, bậc kị bao gồm 2 thuật ngữ và 2 mã nhãn ngữ nghĩa; bậc cụ bao gồm 30 thuật ngữ và 16 mã nhãn ngữ nghĩa; bậc ông bà bao gồm 40 thuật ngữ và 37 mã nhãn ngữ nghĩa; bậc cha mẹ bao gồm 74 thuật ngữ và 20 mã nhãn ngữ nghĩa; bậc bác chú, cô, cậu, gi/ di bao gồm 24 thuật ngữ và 24 mã nhãn ngữ nghĩa; bậc vợ chồng bao gồm 24 thuật ngữ và 11 mã nhãn ngữ nghĩa; bậc anh chị em bao gồm 62 thuật ngữ và 44 mã nhãn ngữ nghĩa; bậc con cháu bao gồm 49 thuật ngữ và 45 mã nhãn ngữ nghĩa (xem Bảng 1). Từ các nhánh chính của hệ thống ý niệm thuật ngữ thân tộc, các chi nhánh con phát triển thêm với độ sâu phù hợp tùy theo đặc điểm của từng nhánh.

Bảng 1. Số lượng thuật ngữ thân tộc và mã nhãn ngữ nghĩa

Bậc	Số lượng thuật ngữ thân tộc	Số lượng mã nhãn ngữ nghĩa
Bậc kị	2	2
Bậc cụ	30	16
Bậc ông bà	40	37
Bậc cha mẹ	74	20
Bậc bác chú cô cậu gi/di	24	24

Bậc vợ chồng	24	11
Bậc anh chị em	62	44
Bậc con cháu	49	45

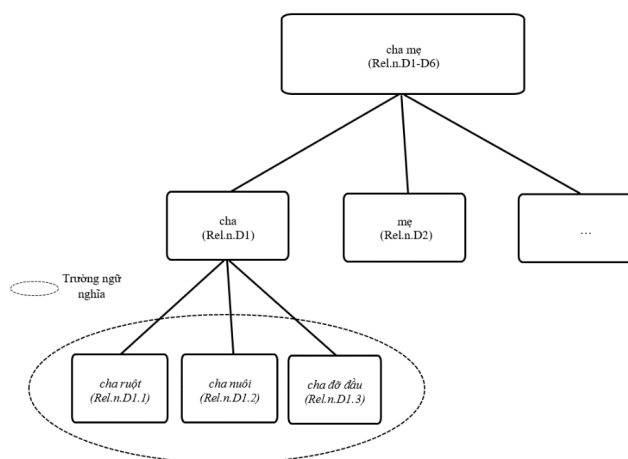
Lấy nhánh thuật ngữ “bậc cha mẹ” (Rel.n.D) làm ví dụ điển hình, hệ thống phân cấp ý niệm được tổ chức như sau: Bắt đầu từ thuật ngữ tổng quát là “bậc cha mẹ” với mã nhãn ngữ nghĩa tổng quát là Rel.n.D, lớp ngữ nghĩa “bậc cha mẹ” (Rel.n.D) này bao gồm lớp kế tiếp “cha mẹ” (Rel.n.D1-D6) và cùng synset với các thuật ngữ đồng nghĩa như: {cha mẹ, bố mẹ, ba mẹ, ba má, phụ mẫu, ông bà già (khẩu ngữ)}. Từ đó, lớp “cha mẹ” được phân chia chi tiết hơn thành 6 nhánh nhỏ khác như: “cha” (Rel.n.D1), “mẹ” (Rel.n.D2), “cha” (Rel.n.D3), “mẹ” (Rel.n.D4), “cha” (Rel.n.D5), “mẹ” (Rel.n.D6). Mỗi nhánh này tiếp tục được phân cấp chi tiết hơn và xuống các tầng sâu hơn. Cụ thể, nhánh “cha” (Rel.n.D1) bao gồm các thuật ngữ như “cha ruột” (Rel.n.D1.1), “cha nuôi” (Rel.n.D1.2), “cha đỡ đầu” (Rel.n.D1.3). Thuật ngữ “mẹ” (Rel.n.D2) cũng tiếp tục phân cấp chi tiết hơn thành các nhánh con như “mẹ ruột” (Rel.n.D2.1), “mẹ nuôi” (Rel.n.D2.2), “mẹ đỡ đầu” (Rel.n.D2.3). Tương tự, “cha” (Rel.n.D3) với nhánh con là “cha dượng” (Rel.n.D3.1), “mẹ” (Rel.n.D4) với nhánh con là “mẹ ghê” (Rel.n.D4.1), “cha” (Rel.n.D5) bao gồm 2 nhánh con là “cha chồng” (Rel.n.D5.1) và “cha vợ” (Rel.n.D5.2), “mẹ” (Rel.n.D6) bao gồm 2 nhánh con là “mẹ chồng” (Rel.n.D6.1) và “mẹ vợ” (Rel.n.D6.2) (xem Bảng 2).

Bảng 2. Minh họa hệ thống nhãn ngữ nghĩa cho “bậc cha mẹ”

STT	MÃ NHÃN NGỮ NGHĨA	THUẬT NGỮ THÂN TỘC
1.	Rel.n.D	Bậc cha mẹ
2.	Rel.n.D1-D6	cha mẹ, bố mẹ, ba mẹ, ba má, phụ mẫu, ông bà già (khẩu ngữ)
3.	Rel.n.D1	cha, bố, ba, thầy, thày, tía, bọ, ông già (khẩu ngữ), cậu (phương ngữ)
4.	Rel.n.D1.1	cha ruột, bố ruột, ba ruột, bố đẻ, thân phụ, phụ thân
5.	Rel.n.D1.2	cha nuôi, bố nuôi, ba nuôi
6.	Rel.n.D1.3	cha đỡ đầu, bố đỡ đầu, ba đỡ đầu
7.	Rel.n.D2	mẹ, má, bầm. bu, u, mế, mạ, bà già (khẩu ngữ), đẻ (phương ngữ Thanh Hóa)
8.	Rel.n.D2.1	mẹ ruột, má ruột, mẹ đẻ, hiền mẫu, thân mẫu, mẫu thân
9.	Rel.n.D2.2	mẹ nuôi, má nuôi
10.	Rel.n.D2.3	mẹ đỡ đầu, má đỡ đầu
11.	Rel.n.D3	cha
12.	Rel.n.D3.1	cha dượng, bố dượng, dượng, cha ghê, kế phụ
13.	Rel.n.D4	mẹ
14.	Rel.n.D4.1	mẹ ghê, dì ghê/ gì ghê, mẹ kế, kế mẫu
15.	Rel.n.D5	cha
16.	Rel.n.D5.1	cha chồng, bố chồng, ba chồng, nhạc phụ
17.	Rel.n.D5.2	cha vợ, bố vợ, ba vợ
18.	Rel.n.D6	mẹ
19.	Rel.n.D6.1	mẹ chồng, má chồng, mạ chồng
20.	Rel.n.D6.2	mẹ vợ, má vợ, mạ vợ, nhạc mẫu

Sự phân chia theo phân cấp ý niệm này phản ánh cách con người liên tưởng đến các từ ngữ trong tâm trí, từ đó hình thành một mạng lưới kết nối các từ vựng về mặt ngữ nghĩa. Các nét nghĩa trong

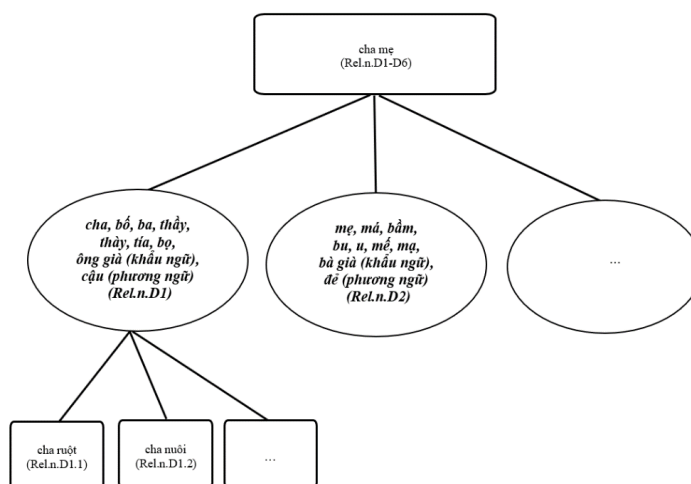
mạng lưới thân tộc này thể hiện mối quan hệ huyết thống hoặc hôn nhân giữa các thành viên trong gia đình, bao gồm cả nét nghĩa nội ngoại và nét nghĩa thể hiện tôn ti thứ bậc. Về mặt cấu trúc, hệ thống này thể hiện sự cân bằng thuộc tính giữa các synsets theo trục ngang, dựa trên cách tổ chức trường ngữ nghĩa với các từ ngữ có cùng cấp độ được phân bổ vào các synset độc lập (xem Hình 2) và mối quan hệ đồng nghĩa (các thuật ngữ đồng nghĩa được đặt trong cùng một synset) (xem Hình 3); đồng thời cấu trúc cũng phân cấp theo trục dọc với mối quan hệ thượng danh và hạ danh (xem Hình 4). Tuy nhiên, cần lưu ý rằng, cấu trúc phân cấp theo mối quan hệ thượng danh và hạ danh trong hệ thống mang nghĩa ý niệm này không giống với nghĩa chủ đề, sẽ không bao hàm nét nghĩa thể hiện tôn ti thứ bậc theo trục dọc.



Hình 2. Minh họa trường ngữ nghĩa cho nhãn ngữ nghĩa “cha” (Rel.n.D1)

4.2. Phân tích các mối quan hệ đồng nghĩa, đa nghĩa, thượng danh và hạ danh

Trong hệ thống thân tộc tiếng Việt, các thuật ngữ được cấu trúc dựa trên các mối quan hệ ngữ nghĩa bao gồm đồng nghĩa (synonymy), đa nghĩa (polysemy), thượng danh (hypernymy) và hạ danh (hyponymy). Từ/ ngữ đồng nghĩa xuất hiện nhiều trong hệ thống này. Các thuật ngữ đồng nghĩa được xếp cùng 1 synset (tập đồng nghĩa) và đặt tên cùng 1 mã nhãn ngữ nghĩa. Chẳng hạn, từ “cha” (Rel.n.D1) có các từ/ ngữ đồng nghĩa như {bố, ba, thầy, thầy, tía, bọ, ông già (khẩu ngữ), cậu (phương ngữ)} hay từ “mẹ” (Rel.n.D1) có các từ/ ngữ đồng nghĩa như {má, bầm, bu, u, mẹ, mạ, bà già, đẽ} (xem Hình 3).

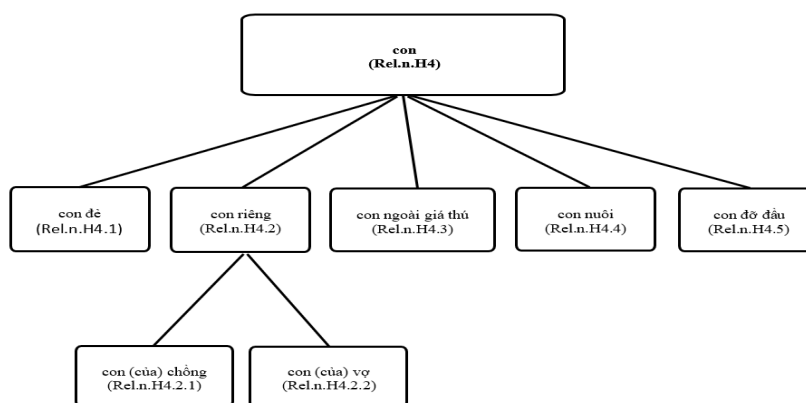


Hình 3. Minh họa mối quan hệ đồng nghĩa của các thuật ngữ thân tộc chỉ “cha” và “mẹ”

Tương tự như trên, thuật ngữ “anh em cọc chèo” ở mã nhãn Rel.n.G3.4 có các thuật ngữ đồng nghĩa như: {*anh em bạn cột chèo, anh em cột chèo, anh em đồng hao, anh em bạn rế, anh em chung số cá, anh em chũn chễn, anh em dùi đục, anh em đứng nằng, anh em sừng bò, anh em tể đũi*}.

Trong hệ thống thân tộc tiếng Việt, chúng tôi nhận thấy sự xuất hiện của mối quan hệ đa nghĩa trong nhiều trường hợp. Một ví dụ điển hình của mối quan hệ đa nghĩa là từ “con” (Rel.n.H1-H5), “con” ở đây có thể chỉ đến “con trai” (Rel.n.H1.1), “con gái” (Rel.n.H1.2), “con dâu” (Rel.n.H1.3), “con rể” (Rel.n.H1.4), “con trưởng” (Rel.n.H2.1), “con thứ” (Rel.n.H2.2), “con út” (Rel.n.H2.3), “con một” (Rel.n.H3.1), “con ruột” (Rel.n.H4.1), “con riêng” (Rel.n.H4.2), “con ngoài giá thú” (Rel.n.H4.3), “con nuôi” (Rel.n.H4.4), “con đỡ đầu” (Rel.n.H4.5), “con so” (Rel.n.H5.1), “con rạ” (Rel.n.H5.2). Tương tự, hệ thống thân tộc còn có các từ/ ngữ đa nghĩa khác như: “ông”, “bà”, “cha”, “mẹ”, “bác”, “chú”, “anh”, “chị”,...

Hệ thống nhãn ngữ nghĩa thân tộc được tổ chức dựa trên mối quan hệ thượng danh và hạ danh, theo chiều dọc, đi từ khái niệm tổng quát đến các khái niệm cụ thể hơn. Chẳng hạn, nhìn từ trên xuống (xem Hình 4), ta thấy “con” (Rel.n.H4) là thượng danh của “con đẻ” (Rel.n.H4.1), “con riêng” (Rel.n.H4.2), “con ngoài giá thú” (Rel.n.H4.3), “con nuôi” (Rel.n.H4.4), “con đỡ đầu” (Rel.n.H4.5). Tiếp tục, “con riêng” là thượng danh của “con (của) chồng” (Rel.n.H4.2.1) và “con (của) vợ” (Rel.n.H4.2.2). Nhìn từ dưới lên, ta thấy “con (của) chồng” là hạ danh của “con riêng”, “con riêng” là hạ danh của “con”. Các thuật ngữ “con đẻ”, “con riêng”, “con ngoài giá thú”, “con nuôi”, “con đỡ đầu” là đồng hạ danh (co-hyponyms) của “con” (Rel.n.H4).



Hình 4. Minh họa mối quan hệ thượng danh và hạ danh cho nhánh thuật ngữ thân tộc “con”

4.3. Ứng dụng của hệ thống nhãn ngữ nghĩa thân tộc tiếng Việt

Hệ thống nhãn ngữ nghĩa cho thuật ngữ thân tộc tiếng Việt có nhiều ứng dụng thực tế trong các lĩnh vực như: giảng dạy tiếng Việt cho người nước ngoài, gán nhãn ngữ nghĩa tự động tiếng Việt, phát triển công cụ học tập và so sánh hệ thống nhãn ngữ nghĩa thân tộc giữa các ngôn ngữ khác nhau.

Đầu tiên, trong lĩnh vực giảng dạy tiếng Việt cho người nước ngoài, hệ thống nhãn ngữ nghĩa thân tộc có thể được sử dụng như một phần tài liệu giảng dạy, giúp người học hiểu rõ hơn về cách xưng hô trong gia đình người Việt và tiếp cận các thuật ngữ thân tộc một cách có hệ thống, từ đó nắm bắt được nét đặc trưng của ngôn ngữ và văn hóa Việt.

Tiếp theo, hệ thống này hỗ trợ gán nhãn ngữ nghĩa tự động, phục vụ cho việc tra cứu nhanh chóng các thuật ngữ thân tộc và xử lý ngôn ngữ tự nhiên. Người dùng có thể dễ dàng tìm kiếm các thuật ngữ liên quan đến thân tộc và khám phá ý nghĩa, cách sử dụng thông qua ứng dụng hoặc nền tảng trực tuyến, giúp tiết kiệm thời gian cho người học. Một ứng dụng quan trọng khác của hệ thống nhãn ngữ nghĩa là khả năng tích hợp vào các công cụ học tập, chẳng hạn như phần mềm học ngôn ngữ, tạo ra các bài tập và trò chơi thú vị giúp người học củng cố kiến thức về thuật ngữ thân tộc một cách sinh động và hiệu quả hơn.

Cuối cùng, việc so sánh các hệ thống nhãn ngữ nghĩa thân tộc giữa các ngôn ngữ khác nhau sẽ cung cấp cơ sở cho các nhà ngôn ngữ học phân tích và so sánh các thuật ngữ thân tộc trong tiếng Việt với các ngôn ngữ khác, giúp người nghiên cứu và người học nhận diện được những điểm tương đồng và dị biệt. Sự khác biệt giữa các hệ thống nhãn ngữ nghĩa thân tộc có thể phản ánh các giá trị văn hóa, lịch sử và xã hội của mỗi quốc gia, từ đó góp phần nâng cao nhận thức về văn hóa và cách duy trì ngôn ngữ của mỗi dân tộc đối với người học.

5. Kết luận

Nghiên cứu này tập trung vào việc xây dựng hệ thống nhãn ngữ nghĩa cho các thuật ngữ thân tộc trong tiếng Việt, với cấu trúc phân cấp đi từ một ý niệm tổng quát đến các ý niệm chi tiết hơn. Thông qua việc phân tích các mối quan hệ ngữ nghĩa như đồng nghĩa, đa nghĩa, thượng danh và hạ danh, hệ thống nhãn ngữ nghĩa này không chỉ cung cấp cách tiếp cận có tổ chức cho các từ ngữ liên quan đến mối quan hệ gia đình mà còn phản ánh giá trị văn hóa đặc trưng trong xã hội Việt Nam. Kết quả đạt được từ nghiên cứu này tạo nền tảng vững chắc cho các nghiên cứu tiếp theo, mở ra cơ hội ứng dụng thực tiễn trong việc giảng dạy tiếng Việt cho người nước ngoài và trong lĩnh vực xử lý tự động ngôn ngữ tự nhiên cũng như trong việc so sánh đối chiếu hệ thống nhãn ngữ nghĩa thân tộc giữa các ngôn ngữ. Mặc dù nghiên cứu đã đem lại hiệu quả tích cực trong việc tổ chức và phân tích hệ thống nhãn ngữ nghĩa, chúng tôi cũng nhận diện được một số thách thức còn tồn tại. Cụ thể, số lượng vốn từ/ ngữ hiện tại có thể chưa bao quát hết toàn bộ các thuật ngữ thân tộc người Việt sử dụng trên các vùng miền khác nhau; đồng thời, nghiên cứu cũng chỉ dừng lại bước đầu ở thuật ngữ thân tộc. Điều này mở ra cơ hội cho các nghiên cứu trong tương lai nhằm bổ sung và hoàn thiện hệ thống nhãn ngữ nghĩa, từ đó phục vụ tốt hơn cho các mục đích giảng dạy, nghiên cứu và phát triển ngôn ngữ.

**Nghiên cứu này được tài trợ bởi Trường Đại học Khoa học Xã hội và Nhân văn, Đại học Quốc gia Thành phố Hồ Chí Minh trong khuôn khổ Đề tài mã số SDH2024 - 04.*

TÀI LIỆU THAM KHẢO

Tiếng Việt

1. Lê Quang Thiêm (2018), *Sự phát triển nghĩa từ vựng tiếng Việt*. Nxb Đại học Quốc gia Hà Nội.
2. Trương Hoàng Bảo, Nguyễn Ngọc Bình, Lê Thành Nguyên và các tác giả khác (2021), Xây dựng WordNet tiếng Việt tự động bằng ngữ liệu song ngữ. *Kỷ yếu Hội nghị KHCN Quốc gia lần thứ XIV về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR)*, Trường Đại học Khoa học Tự nhiên, ĐHQG TP.HCM, ngày 23-24/12/2021, 483-490.
3. Dương Kỳ Đức và Vũ Quang Hào (1994), *Từ điển trái nghĩa - đồng nghĩa tiếng Việt*. Nxb Giáo dục.
4. Dương Thị Nụ (2003), Đối chiếu đặc trưng ngữ nghĩa của nhóm từ chỉ quan hệ thân tộc trong tiếng Anh và tiếng Việt. *Đề tài nghiên cứu khoa học cấp đại học quốc gia, Trường Đại học Ngoại ngữ, ĐHQG Hà Nội*.
5. Đặng Thanh Hòa (2005), *Từ điển phương ngữ tiếng Việt*. Nxb Đà Nẵng.
6. Hoàng Phê (Chủ biên) (tái bản năm 2020), *Từ điển tiếng Việt*. Nxb Đà Nẵng - Trung tâm Từ điển học, Hà Nội.
7. Huỳnh Công Tín (2007), *Từ điển Từ ngữ Nam Bộ*. Nxb Khoa học xã hội.
8. Nguyễn Như Ý, Đặng Ngọc Lệ và Phan Xuân Thành (2001), *Từ điển đối chiếu Từ địa phương*. Nxb Giáo dục.
9. Nguyễn Văn Khang (2014), “Giao tiếp xung hô tiếng Việt bằng từ thân tộc và việc sử dụng chúng trong giao tiếp công quyền”. *Tạp chí Ngôn ngữ & Đời sống*, Số 10, 38-46.
10. Nguyễn Võ Tâm Như (2023), *Đối chiếu từ ngữ chỉ quan hệ thân tộc trong tiếng Việt và tiếng Nhật từ lý thuyết lược đồ vai trò*. Luận văn Thạc sĩ chuyên ngành Ngôn ngữ học, mã số: 8229020, Trường Đại học Sài Gòn.

Tiếng Anh

11. Fellbaum, C. (Ed.). (1998), *WordNet: An electronic lexical database*. MIT press.
12. Fellbaum, C. (2005), WordNet and wordnets. In: Brown, Keith et al. (eds.), *Encyclopedia of Language and Linguistics, Second Edition*, Oxford: Elsevier, 665-670. Retrieved from <https://wordnet.princeton.edu/>
13. Lam, K. N. & Kalita, J. (2022), Constructing Vietnamese wordnet: A case study. *Computación y Sistemas*, 26(3), 1313-1322.
14. Leech, G. (1985), *Semantics: The study of meaning*. England: Penguin Books Ltd.
15. McArthur, T. (1988), *Longman Lexicon of Contemporary English*. London: Longman.
16. Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International journal of lexicography*, 3(4), 235-244.
17. Rayson, P., Archer, D., Piao, S., & McEnery, A. M. (2004), The UCREL semantic analysis system. Retrieved from https://www.lancaster.ac.uk/staff/rayson/publications/usas_lrec04ws.pdf
18. Vossen, P. (2002), Wordnet, Eurowordnet and global wordnet. *Revue française de linguistique appliquée*, 7(1), 27-38.

NGỮ LIỆU

- [1] Phạm Văn Hảo (2009), *Từ điển phương ngữ tiếng Việt*. Nhà xuất bản Khoa học xã hội.
- [12] Các tác phẩm văn học (truyện ngắn, tiểu thuyết), báo chí và các tài liệu chính thống khác của Việt Nam.

Semantic tagset system of Vietnamese family relationship terms

Abstract: Family relationships are a common topic in the study of language and culture. However, there is currently no semantic tagset system for family relationship terms in Vietnamese from a linguistic perspective. This research employed descriptive statistics and analysis-synthesis methods to describe, classify, and systematize Vietnamese family relationship terms according to a hierarchical conceptual structure and semantic relationships such as synonymy, polysemy, hypernymy, and hyponymy. The results of this study open up potential applications in teaching Vietnamese to foreigners, automatic semantic processing, and comparative semantics research.

Key words: Vietnamese family relationship terms; semantic tagset system; synonymy; polysemy; hierarchical conceptual structure.