

KẾT NGỮ TRONG SÁCH CHUYÊN NGÀNH TIẾNG ANH CÔNG TRÌNH THỦY COLLOCATIONS IN ENGLISH PROFESSIONAL BOOKS OF HYDRAULIC ENGINEERING

NGUYỄN HỒNG ÁNH^{1*}, VŨ THỊ THÚY¹, LÊ THỊ HƯƠNG GIANG²

Khoa Ngoại ngữ, Trường Đại học Hàng hải Việt Nam

Khoa Công trình, Trường Đại học Hàng hải Việt Nam

*Email liên hệ: anhh.nn@vimaru.edu.vn

Tóm tắt

Bài báo này tường thuật công trình nghiên cứu thử nghiệm về việc xây dựng danh mục kết ngữ trong các văn bản chuyên ngành bằng tiếng Anh nhằm phục vụ công tác giảng dạy chuyên ngành bằng tiếng Anh và biên soạn từ điển kết ngữ chuyên ngành. Đề tài lựa chọn thí điểm các sách chuyên ngành bằng tiếng Anh phục vụ các học phần chuyên ngành Công trình thủy tại Trường Đại học Hàng hải Việt Nam. Quá trình nghiên cứu bắt đầu bằng việc chiết xuất thuật ngữ chuyên ngành Công trình thủy, sau đó tìm kiếm các kết ngữ phát sinh trong khoảng tìm kiếm ± 3 . Phương pháp nghiên cứu kết hợp phương pháp định lượng (thống kê, so sánh tần số) và định tính (phân tích, kiểm định kết cấu và ngữ nghĩa) đối với từng thuật ngữ và kết ngữ. Kết quả thu được gồm 21.760 kết ngữ (2-5 từ) có tần số xuất hiện cao trong ngữ liệu ban đầu với 3,06 triệu từ.

Từ khóa: Kết ngữ, thuật ngữ chuyên ngành, sách chuyên ngành, Công trình thủy.

Abstract

This article reports a pilot study on developing a list of specialized collocations in English professional books to serve the English-Medium Instruction at VMU and the compilation of specialized collocation dictionaries. The study collected specialized books in Hydraulic engineering to build specialized corpus. The collocation extraction started with the identification of specialised terminologies in Hydraulic engineering, then identified their collocates within a ± 3 span. The extraction method combined a quantitative approach (calculating and comparing the frequency of occurrence for each terminology and collocation) and qualitative approach (analyzing syntactic structure and semantics) for every terminologies and collocation. The final result displayed 21,760 collocations (2-5 words) of high frequency from the original 3,06-million-word corpus.

Keywords: Lexical collocations, specialised terminologies, professional books, quaywall engineering.

1. Đặt vấn đề

Nghiên cứu đương đại về ngôn ngữ đã chứng minh được vai trò trung tâm của từ vựng đối với khả năng thụ đắc ngôn ngữ. Từ vựng là nền móng cơ sở hình thành nên ngôn ngữ, và không có từ vựng thì không có ngôn ngữ. Quan điểm truyền thống cho rằng khi người học ngoại ngữ thuộc được càng nhiều từ, họ càng có khả năng vận dụng từ phong phú hơn trong giao tiếp. Nhưng thực tế cho thấy để đạt được khả năng sử dụng ngoại ngữ lưu loát, việc học thuộc các từ đơn lẻ là chưa đủ, người học cần phải biết cách kết hợp các từ theo chuỗi đặc thù [1]. Các chuỗi từ có sẵn, mang tính chất lặp đi lặp lại chiếm tỷ trọng lớn trong hầu hết văn bản hoặc ngôn bản.

Kết ngữ (collocations), một thể loại đặc trưng trong các dạng chuỗi từ, có vai trò quan trọng giúp người học ngoại ngữ sử dụng ngoại ngữ trôi chảy, tự nhiên gần như người bản ngữ. Việc sử dụng thành thạo kết ngữ là một trong những cách chuẩn xác giúp phân biệt được người bản ngữ và người nước ngoài. Người bản ngữ có khả năng vận dụng kết ngữ một cách tự động, khiến họ giao tiếp trôi chảy vì các kết ngữ là chuỗi từ đã có sẵn trong tư duy của họ. Khi người học ngoại ngữ tích lũy đủ vốn từ vựng liên quan đến kết ngữ, họ dễ dàng tránh được các lỗi dùng sai từ khi nói hoặc viết.

Tuy nhiên, đối với người học ngoại ngữ nói chung và tiếng Anh nói riêng, kết ngữ rất khó nhận diện và sử dụng đúng cách. Trong kết ngữ luôn có một thành phần không dễ dự đoán và không phải lúc nào cũng có thể thay thế bằng một từ khác có nghĩa tương đương (VD: các từ "rotten", "addled", "sour" và "rancid" có nghĩa tương tự nhưng người Anh chỉ nói *rotten egg*, *addled egg*, *sour milk*, và *rancid butter*). Cho dù người học đã học tiếng Anh bao lâu đi chăng nữa, trình độ tiếng Anh của họ tốt đến mức nào đi nữa, và mức tương đồng giữa tiếng Anh và tiếng mẹ đẻ cao đến thế nào đi nữa, họ vẫn có thể mắc lỗi về kết ngữ do chịu ảnh hưởng từ tiếng mẹ đẻ [2]. Các lỗi sai liên quan đến kết ngữ này khiến cách hành văn khi nói và viết tiếng Anh bất thường và lạ kiêu. Nghiên cứu của Bahns và Eldaw [3] cho thấy người dịch thuật mắc lỗi dịch sai kết ngữ cao gấp đôi lỗi dịch sai từ đơn.

Thật không may, kết ngữ chiếm một tỷ trọng rất lớn trong từ vựng tiếng Anh. Hầu như danh từ, tính từ, động từ nào trong tiếng Anh cũng có kết ngữ. Trong các văn bản chuyên ngành bằng tiếng Anh, kết ngữ được sử dụng lặp lại với tần suất cao, chiếm 70% nội dung nghe, nói, đọc viết [4]. Chính vì vậy, người sử dụng tiếng Anh phải học thuộc hoặc phải tra cứu kết ngữ trong từ điển thích hợp. Việc biên soạn các từ điển kết ngữ trở nên vô cùng quan trọng, giúp người dùng không chỉ tra cứu thuật ngữ chuyên ngành mà còn cả các kết ngữ kèm theo [5].

Sự phát triển về công nghệ máy tính đã cho phép các nhà nghiên cứu từ vựng xây dựng ngữ liệu chuyên ngành và dựa vào các phép tính thống kê để trích xuất thuật ngữ chuyên ngành và kết ngữ thay vì chỉ dựa vào trực giác của họ. Trong thực tế, hiếm có kết ngữ nào dài hơn 6 từ, nên trong mỗi một tổ hợp từ (word combination), thành phần kết hợp (collocate) được xác định là từ xuất hiện ở bên trái hoặc bên phải của thuật ngữ (terminology) trong khoảng ± 3 (span). Nếu từ kết hợp xuất hiện với tần suất cao, có độ liên kết mạnh với thuật ngữ (chỉ số tương hỗ MI ≥ 3), xác suất kết hợp cao (giá trị kiểm định $t \geq 2$), thì sự kết hợp giữa chúng mới được coi là kết ngữ [6].

Vì đây là công trình thí điểm, đề tài chỉ tập trung vào các kết ngữ từ vựng (lexical collocations). Các từ khoá chuyên ngành và từ kết hợp với chúng để tạo kết ngữ thuộc các đơn vị từ loại sau: danh từ, động từ, tính từ và trạng từ. Đề tài bỏ qua kết ngữ ngữ pháp (grammatical collocations), loại tổ hợp từ chứa từ chức năng (VD: giới từ, liên từ, từ hạn định), vì loại tổ hợp này có thể được tra cứu trong các từ điển tiếng Anh phổ thông.

2. Phương pháp nghiên cứu

Quy trình chiết xuất và lọc kết ngữ gồm các bước sau: (1) Xây dựng ngữ liệu, (2) Xác định thuật ngữ chuyên ngành (terminologies), (3) Xác định và lọc kết ngữ, và (4) Hệ thống hóa kết ngữ. Quy trình chiết xuất tự động sử dụng các phép đo định lượng về tần suất, tỷ lệ phân bố, giá trị MI, giá trị t của từng tổ hợp từ với công cụ Sketch Engine [7] và WordSmith Tools 7.0 [8]. Quy trình lọc giúp lựa chọn các kết ngữ có giá trị sử dụng thực tế cao.

2.1. Xây dựng ngữ liệu

a. *Ngữ liệu chuyên ngành (Specialised corpus)*: trong khuôn khổ của đề tài, 34 đầu sách chuyên ngành Công trình thủy (xem Phụ lục) được sử dụng để xây dựng ngữ liệu chuyên ngành với tổng số 3,06 triệu từ bao phân bố trên 9 nội dung chuyên ngành: (1) Tin học ứng dụng (Applied software), (2) Động lực học sông biển (River and Sea Dynamics), (3) Khí tượng, thủy hải văn (Hydrology and Meteorology), (4) Quy hoạch cảng (port planning), (5) Công trình bến & công trình cảng (Port and Harbour structures), (6) Đê chắn sóng (Breakwaters), (7) Công trình bảo vệ bờ (Shore protection structures), (8) Công trình ven biển và ngoài khơi (Coastal and Offshore structures), và (9) Công trình thủy công (service systems in ports). Việc gắn nhãn từ loại (POS-tagging) được áp dụng trong toàn ngữ liệu với CLAWS5 [9].

b. *Ngữ liệu tham chiếu (Reference corpus)*: ngữ liệu tham chiếu là tập hợp tài liệu tiếng Anh tổng quát (dạng nói và viết) Open American National Corpus [10]. Ngữ liệu tham chiếu được sử dụng để so sánh đối chiếu với ngữ liệu chuyên ngành, giúp nhóm tác giả nghiên cứu xác định được thuật ngữ chuyên ngành Công trình thủy. Với tổng lượng từ là 15 triệu từ, ngữ liệu tham chiếu trong đề tài này phù hợp tiêu chuẩn quy ước: lớn gấp 5 lần kích cỡ ngữ liệu chuyên ngành.

Bảng 1. Danh mục 34 sách chuyên ngành Công trình thủy

| TT | Danh mục sách chuyên ngành Công trình thủy |
|----|--|
| 1 | Mathsoft, Inc. (2001). <i>Mathcad: user's guide with reference manual: Mathcad 2001 professional</i> . MathSoft, Inc.. |
| 2 | SAP2000, C. S. I. (2011). <i>Linear and nonlinear static and dynamic analysis and design of three-dimensional structures. Computers and Structures. Inc.: Berkeley, California, USA</i> . |
| 3 | Danish Hydraulic Institute (DHI). (2007). MIKE 21 flow model: Hydrodynamic module. User guide. <i>Danish Hydraulic Institute Water and Environment</i> , 1-89. |
| 4 | Stolarski, T., Nakasone, Y., & Yoshimoto, S. (2018). <i>Engineering analysis with ANSYS software</i> . Butterworth-Heinemann. |
| 5 | Chanson, H. (2004). <i>Hydraulics of open channel flow</i> . Elsevier. |
| 6 | Montañés, J. L. (2005). <i>Hydraulic canals: design, construction, regulation and maintenance</i> . CRC Press. |
| 7 | Graebel, W. (2007). <i>Advanced fluid mechanics</i> . Academic Press. |
| 8 | Annandale, G. W. (2006). <i>Reservoir sedimentation</i> . Encyclopedia of Hydrological Sciences. |
| 9 | Joseph, P., & Lomas, S. A. (2004). <i>Deep-water sedimentation in the Alpine basin of SE France: New perspectives on the Grés d'Annot and related systems</i> . Geological Society of London. ISBN electronic: 9781862394698 |
| 10 | Levec, F., & Skinner, A. (2004). <i>Manual of instructions: Bathymetric surveys</i> . Ontario Ministry of Natural Resources. |
| 11 | IHO, C. (2005). <i>Manual of Hydrography</i> . Monaco: International Hydrographic Bureau |
| 12 | Sene, K. (2015). <i>Hydrometeorology: forecasting and applications</i> . Springer. |

| | |
|----|--|
| 13 | Davie, T. (2008). <i>Fundamentals of hydrology</i> . Routledge. |
| 14 | HEADQUARTERS, U. A. C. O. E. (1989). <i>Water levels and wave heights for coastal engineering design</i> . Washington, DC: Engineer Manual EM, 1110-2. |
| 15 | Tsinker, G. P. (Ed.). (2004). <i>Port engineering: planning, construction, maintenance, and security</i> . John Wiley & Sons. |
| 16 | Groenveld, R. (1993). <i>Service systems in ports and inland waterways</i> . TU. |
| 17 | Bruun, P. (2006). <i>Port Engineering: Harbor planning, breakwaters, and marine terminals</i> . Gulf Publishing Company |
| 18 | Thoresen, C. A. (2010). <i>Port designer's handbook</i> . Thomas Telford Ltd. |
| 19 | Ligteringen, H. (2017). <i>Ports and Terminals</i> . Delft Academic Press |
| 20 | Port, J. (1999). Technical standards and commentaries for port and harbour facilities. <i>The Japan Port & Harbour Association, 1</i> , 455-473. |
| 21 | McBride, M., Boll, M., & Briggs, M. (2014). Harbour approach channels—Design guidelines. <i>PIANC Report No. 121</i> . |
| 22 | Gaythwaite, J. W. (2016, November). <i>Design of marine facilities: engineering for port and harbor structures</i> . American Society of Civil Engineers. |
| 23 | Goda, Y (2000). <i>Random sea and design of marine structures</i> . Singapore: World Scientific Publishing Co. Pte. Ltd. |
| 24 | Gerwick Jr, B. C. (2007). <i>Construction of marine and offshore structures</i> . CRC press. |
| 25 | Sharkov, E. A. (2007). <i>Breaking ocean waves: geometry, structure and remote sensing</i> . Springer Science & Business Media. |
| 26 | d'Angremond, K., & Van Roode, F. (2018). <i>Breakwaters and closure dams</i> . CRC Press. |
| 27 | Kim, Y. C. (2015). Design of Coastal Structures and Sea Defenses. <i>Design of Coastal Structures and Sea Defenses</i> . Singapore: World Scientific Publishing Co. Pte. Ltd. |
| 28 | Schiereck, G. J. (2017). <i>Introduction to bed, bank and shore protection</i> . CRC Press. |
| 29 | Goudas, C., Katsiaris, G., May, V., & Karambas, T. (Eds.). (2003). <i>Soft shore protection: An environmental innovation in coastal engineering</i> (Vol. 7). Springer Science & Business Media. |
| 30 | Przedwojski, B., Błażejowski, R., & Pilarczyk, K. W. (1995). <i>River training techniques: fundamentals, design and applications</i> . AA Balkema. |
| 31 | McCartney, B. L., Lee, B. K., George, J., Neilson, F., & Lindgren, M. (1998). <i>Inland navigation: Locks, dams, and channels</i> . American Society of Civil Engineers. |
| 32 | Glerum, A., & Vrijburgh, A. (2000). Design of locks 1. <i>Rijkswaterstaat</i> . |
| 33 | U.S. Army Corps of Engineers (1995). <i>Construction Control for Earth and Rock-Fill Dams</i> . Washington, DC: Engineer Manual EM 1110-2-1911. |
| 34 | Yazdandoost, F., & Attari, J. (2004). <i>Hydraulics of Dams and River Structures</i> . CRC Press. |

2.2. Trích xuất thuật ngữ chuyên ngành

Thuật ngữ chuyên ngành trong đề tài này được xác định là các từ đặc trưng cho chuyên ngành công trình thủy, tức là tỷ lệ xuất hiện trong ngữ liệu chuyên ngành cao vượt trội so với trong ngữ liệu tham chiếu ($p < 0,000001$ (sai số 1/1 triệu)). Ngoài ra trong ngữ liệu tham chiếu, các tiêu chí thống kê khác bao gồm: tần số tối thiểu là 15 (~5 lần/ 1 triệu từ), trị số thuật ngữ (keyness) tối thiểu 1,5.

2.3. Xác định kết ngữ

Cân nhắc đối tượng phục vụ bao gồm sinh viên đại học ngành Công trình thủy (trình độ tiếng Anh chủ yếu ở ngưỡng A2-B1), đề tài hướng đến các kết ngữ phổ biến nhất (đối với kết ngữ gồm 2 từ thì tần số tối thiểu là 3 lần/ 1.000.000 từ. Vì thế, tần số tối thiểu 9 lần trên toàn ngữ liệu được áp dụng đối với kết ngữ 2 từ. Các kết ngữ dài hơn thường có số lượng ít hơn, nên nhóm nghiên cứu không áp đặt mức tần số tối thiểu để tránh bỏ sót các cụm từ đặc thù chuyên ngành hẹp. Tuy nhiên để được công nhận là kết ngữ, tất cả các tổ hợp từ dù ngắn hay dài phải đạt điều kiện $MI \geq 3$, và $t \geq 2$.

Việc kiểm định ngữ cảnh sử dụng của tổ hợp từ giúp loại trừ các tổ hợp từ không hữu ích về mặt sự phạm do không hoàn chỉnh về cú pháp hoặc ngữ nghĩa (Ví dụ: *passive earth*).

2.4. Hệ thống hóa kết ngữ

Việc hệ thống hóa được thực hiện theo quy ước sắp xếp từ thông thường trong từ điển chuyên ngành. Các thuật ngữ dạng nguyên thể được sử dụng làm từ tra cứu. Các từ kết hợp được phân chia theo đơn vị từ loại. Ví dụ, với thuật ngữ là danh từ (hoặc cụm danh từ), nhóm từ kết hợp đầu tiên là các động từ lấy thuật ngữ đó làm chủ ngữ. Nhóm thứ 2 là các động từ lấy thuật ngữ làm tân ngữ. Nhóm thứ 3 là các tính từ bổ nghĩa. Nhóm thứ 4 là các danh từ bổ nghĩa. Trong quá trình này, các kết ngữ mở rộng có chứa thêm giới từ sẽ được bổ sung nếu chúng xuất hiện quá thường xuyên, do đó có giá trị sự phạm cao. Ví dụ: *E.g. bear resemblance (to), determined based on/upon*.

3. Kết quả phân tích và thảo luận

Kết quả phân tích, đối chiếu giữa ngữ liệu chuyên ngành và ngữ liệu tham chiếu thu được 3.117 thuật ngữ trong đó có 1.755 từ đơn và 1.362 từ ghép gồm hai từ trở lên. Thống kê cho thấy hầu như toàn bộ các thuật ngữ là từ đơn hoặc từ ghép gồm 2-3 từ (98,94%), ngược lại số lượng

thuật ngữ gồm 4 từ trở lên chiếm tỷ trọng vô cùng ít ỏi (1,06%). Bảng 2 biểu thị 10 từ đơn và 10 từ ghép có tần số cao nhất.

Bảng 2. Các thuật ngữ phổ biến nhất trong các tài liệu Công trình thủy

| Thuật ngữ (từ đơn) | | | | | Thuật ngữ (từ ghép) | | | | |
|--------------------|------------|---|---------------------------------------|-------------------|---------------------|--------------------|---|---------------------------------------|-------------------|
| | Thuật ngữ | Tần số (ngữ liệu chuyên ngành) | Tần số (ngữ liệu tham chiếu) | Trị số keyness | | Thuật ngữ | Tần số (ngữ liệu chuyên ngành) | Tần số (ngữ liệu tham chiếu) | Trị số keyness |
| 1 | wave | 17.468 | 8.234 | 33,53 | 1 | earth pressure | 1.151 | 0 | 4,77 |
| 2 | pile | 7.897 | 3.188 | 20,90 | 2 | water level | 1.093 | 158 | 4,51 |
| 3 | load | 5.024 | 5.887 | 11,43 | 3 | sheet pile | 836 | 0 | 3,73 |
| 4 | soil | 4.601 | 4.987 | 11,11 | 4 | water depth | 804 | 15 | 3,63 |
| 5 | flow | 5.282 | 7.835 | 10,76 | 5 | sheet piling | 597 | 4 | 2,95 |
| 6 | equation | 4.004 | 3.516 | 10,73 | 6 | shallow water | 582 | 117 | 2,87 |
| 7 | depth | 3.784 | 2.990 | 10,56 | 7 | water pressure | 532 | 33 | 2,73 |
| 8 | concrete | 3.650 | 2.644 | 10,47 | 8 | sediment transport | 495 | 2 | 2,62 |
| 9 | breakwater | 2.865 | 82 | 10,30 | 9 | bearing capacity | 446 | 4 | 2,46 |
| 10 | structure | 8.018 | 18.801 | 10,18 | 10 | deep water | 450 | 165 | 2,44 |

Từ 3.117 thuật ngữ chuyên ngành, việc tìm kiếm từ kết hợp về bên trái hoặc bên phải thuật ngữ (trong khoảng ± 3) đã thống kê được 21.760 kết ngữ hoàn chỉnh:

- 10 kết ngữ hai từ phổ biến nhất là: *wave height* (2349), *sheet pile* (958), *water depth* (864), *pile wall* (416), *wave period* (408), *wave action* (382), *wave overtopping* (357), *wave condition* (273), *wave force* (372), *reinforced concrete* (367).

- 10 kết ngữ ba từ phổ biến nhất là: *passive earth pressure* (384), *sheet pile wall* (387), *active earth pressure* (367 lần), *steel sheet piling* (181), *water pressure difference* (166), *mean water level* (118), *flow (of) sediment transport* (76), *front sheet piling* (46), *shallow water wave* (37), và *sheet piling wall* (36).

- 10 kết ngữ bốn từ phổ biến nhất là: *active earth pressure side* (18), *active earth pressure line* (12), *determine (the) active earth pressure side* (11), *angle (of) active earth pressure* (9), *passive earth pressure side* (16), *calculate (the) passive earth pressure* (11), *passive earth pressure area* (10), *mobilized passive earth pressure side* (10), *joint probability density function* (14), và *water pressure difference + act* (10).

- 10 kết ngữ năm từ phổ biến nhất là: *steel sheet pile cellular-bulkhead quaywall* (14), *executed example of repairing damage* (10), *active earth pressure failure plane* (8), *passive earth pressure failure plane* (6), *applicable (to) earth pressure failure plane* (6), *standard allowable axial bearing capacity* (7), *maximum size of coarse aggregate* (7), *active earth pressure line + draw* (6), *characteristic passive earth pressure force* (6), *follow (a) limit state design method* (6).

- Không tồn tại kết ngữ sáu từ vì tần số quá ít (1-2 lần) và giá trị MI và giá trị t không thỏa mãn yêu cầu.

Về mặt kết cấu, 21.760 kết ngữ từ vựng được phân thành 5 nhóm với số lượng cụ thể như Bảng 3.

Bảng 3. Thông tin khái quát về kết ngữ chuyên ngành Công trình thủy

| Nhóm | Mẫu | Kết ngữ 2 từ | | Kết ngữ 3 từ | | Kết ngữ 4 từ | | Kết ngữ 5 từ | |
|------|---|---------------|--------|--------------|--------|--------------|--------|--------------|--------|
| T1 | <ul style="list-style-type: none"> • A + N(P) • N + N(P) • N(P) + N • Quan + N(P) | 11.809 | 70,14% | 3311 | 75,27% | 376 | 80,17% | 47 | 81,03% |
| T2 | <ul style="list-style-type: none"> • V + N(P) • V + PrepP | 1.867 | 11,09% | 487 | 11,07% | 54 | 11,51% | 7 | 12,07% |
| T3 | <ul style="list-style-type: none"> • N(P) + V | 1.539 | 9,14% | 433 | 10,06% | 39 | 8,32% | 4 | 6,90% |
| T4 | <ul style="list-style-type: none"> • V + Adv • Adv + V | 1.271 | 7,55% | 108 | 2,46% | 0 | 0% | 0 | 0% |
| T5 | <ul style="list-style-type: none"> • Adv + A | 350 | 2,08% | 50 | 1,14% | 0 | 0% | 0 | 0% |
| | Tổng | 16.834 | | 4.399 | | 469 | | 58 | |

Ví dụ: Với thuật ngữ “earth pressure”, kết quả thống kê phát hiện 759 kết ngữ (3 từ) phù hợp các điều kiện nêu ra và các kết ngữ này được phân bổ vào 3 nhóm T1, T2, T3 như trong Bảng 4.

Bảng 4. Kết ngữ với thuật ngữ “earth pressure”

| Nhóm | Mẫu | Ví dụ | Số lượng | Tỉ lệ |
|------|---|--|----------|-------|
| T1 | <ul style="list-style-type: none"> • A + N(P) • N + N(P) • Quan + N(P) | - passive + earth pressure - active + earth pressure - lateral + earth pressure - resultant + earth pressure - | 592 | 78% |
| T2 | <ul style="list-style-type: none"> • V + N(P) • V + PrepP | - calculate + (the) earth pressure - determine + (the) earth pressure - increase + (the) earth pressure - reduce + (the) earth pressure - mobilise + (the) earth pressure - ... | 129 | 17% |
| T3 | <ul style="list-style-type: none"> • N(P) + V | - earth pressure + act - earth pressure + slide - earth pressure + shield - ... | 38 | 5% |

Hình 1 sau đây hiển thị ngữ cảnh sử dụng kết ngữ “calculate+ (the) earth pressure” trong đó “earth pressure” làm tân ngữ cho động từ “calculate”.

| Details | Left context | KWIC | Right context |
|--|-------------------------------------|------------------|--|
| 1 ⓘ doc#1 illy incorrect to | calculate the earth | pressure | using full wall friction on a cantilever |
| 2 ⓘ doc#9 n neglected . 4 . | The seismic earth | pressures | are calculated using the most wide |
| 3 ⓘ doc#24 14 . 2 . 5) (14 . 2 . 6) | Active earth | pressure | is calculated using equation (14 . 2 . 6) |
| 4 ⓘ doc#24 uld be used to | calculate the earth | pressure | of cohesive soil down to the sea bot |
| 5 ⓘ doc#24 s zero when | calculating the earth | pressure | at the depth of 10 m from the sea b |
| 6 ⓘ doc#24 be standard to | calculate the earth | pressure | acting below the residual water leve |
| 7 ⓘ doc#24 ess of soil layer to | calculate earth | pressure | below the residual watere level (m |
| 8 ⓘ doc#24 gnored when | calculating the earth | pressure | of filling , wave force , uplift pressur |
| 9 ⓘ doc#24 ility of Slopes . To | calculate earth | pressure | on sheet piles or anchorage inside t |
| 10 ⓘ doc#24 method . When | calculating earth | pressure | of lightweight treated soil under ordi |

Hình 1. Kết ngữ “calculate + earth pressure” trong ngữ cảnh “động từ + tân ngữ”

Thông tin từ Bảng 3 (và ví dụ từ Bảng 4) cho thấy sự “thống trị” của kết ngữ với danh từ. Nguyên nhân chủ yếu là do (cụm) danh từ chiếm số lượng chủ đạo trong danh mục thuật ngữ, dẫn đến sự vượt trội về số lượng của kết ngữ với (cụm) danh từ. Ngoài ra về bản chất, danh từ cũng có khuynh hướng kết hợp với các đơn vị từ loại nhiều hơn so với các lớp từ khác. Sự thống trị của kết ngữ với danh từ ở đây phản ánh đặc trưng của loại hình văn bản học thuật (cụ thể ở đây là sách chuyên ngành): sự danh từ hoá chiếm tỷ lệ rất lớn. Hiện tượng danh từ hóa rất phổ biến trong các tài liệu học thuật hay văn bản chuyên ngành trong việc truyền đạt nhiều nội dung thông tin một lúc, điều này gây ra những khó khăn nhất định cho việc đọc hiểu văn bản [11].

Các kết ngữ với động từ xếp hạng thứ hai về số lượng nhưng chúng chỉ chiếm một tỷ lệ nhỏ so với các kết ngữ với danh từ. Tuy nhiên, việc dạy và học cũng nên chú trọng loại hình kết ngữ với động từ. Trong những năm gần đây, kết ngữ với động từ đã nhận được nhiều sự quan tâm hơn so với trước kia, bởi vì các công trình nghiên cứu đã phát hiện người học ngoại ngữ thường gặp nhiều khó khăn với các kết cấu V+N(P) hơn các loại kết cấu khác [12].

Ngoài các phát hiện nói trên, vai trò của việc kiểm tra và tinh lọc thủ công đối với kết ngữ cần được thảo luận ở đây. Thống kê định lượng qua phần mềm máy tính đã hỗ trợ rất lớn trong việc tìm kiếm kết ngữ, nhưng vẫn chưa đủ để chiết xuất ra danh mục kết ngữ có giá trị sử dụng cao. Nguyên nhân thứ nhất là các thuật ngữ được trích xuất tự động không được gộp theo gốc từ, từ loại và hậu tố, dẫn đến các biến thể của động từ (thì thời, danh động từ, phân từ), danh từ (số ít, số nhiều), tính từ (cấp so sánh) được xử lý như các từ riêng biệt. Việc tinh giản là cần thiết để tránh cung cấp quá nhiều thông tin gây quá tải cho người học. Để giảm bớt số lượng kết ngữ lặp, việc kiểm tra một cách thủ công phải được tiến hành với từng cụm riêng lẻ.

Nguyên nhân thứ hai là do một số kết ngữ đã bị máy tính loại trừ do tần số xuất hiện không cao. Thực tế là từng biến thể của kết ngữ (do thay đổi về hậu tố -s/-es/-ed/-ing) không đáp ứng yêu cầu về tần số tối thiểu, nhưng tổng số lần xuất hiện của các biến thể lại thực sự cao hơn ngưỡng tần số quy định. Điều tương tự đã xảy ra trong nghiên cứu của Ackermann and Chen [13], khi nhóm chuyên gia phát hiện một số kết ngữ học thuật cần dùng lại bị bỏ sót trong quá trình thống kê trên máy tính và rồi họ đã quyết định dò tìm lại các cụm bị bỏ sót đó. Trong quá trình lọc, họ đã bổ sung thêm 239 kết ngữ (9,7%) vào danh mục sử dụng. Trong công trình nghiên cứu này, nhóm nghiên cứu áp dụng thao tác tương tự và đã bổ sung được 508 kết ngữ (3,02%) vào danh mục.

Từ các thao tác thủ công trên, có thể khẳng định rằng quá trình tinh lọc thủ công mặc dù được diễn giải với các bước vô cùng ngắn gọn, nhưng việc thực hiện rất tốn thời gian và thực sự có tác dụng rất lớn trong việc nâng cao chất lượng của danh mục kết ngữ. Do đó, để phát triển danh mục kết ngữ, không nên chỉ đơn thuần ý lại hay phụ thuộc hoàn toàn vào các kết quả phân tích tính toán từ máy tính.

4. Kết luận

NNghiên cứu đã xây dựng được danh mục 21.760 kết ngữ trích xuất từ 3.117 thuật ngữ chuyên ngành Công trình thủy. Danh mục này là một công cụ hữu ích đối với giảng viên dạy ngoại ngữ trong việc xây dựng các tài liệu giảng dạy phù hợp và định hướng sinh viên tập trung vào các mục kết ngữ thường gặp thay vì các từ đơn hoặc thuật ngữ riêng lẻ. Từ danh mục này, các chuyên gia từ vựng học có thể biên soạn từ điển chuyên ngành phục vụ giảng viên, sinh viên, kỹ sư và các nhà nghiên cứu trong lĩnh vực công trình thủy. Việc biên soạn nên có sự phối hợp với các chuyên gia trong lĩnh vực chuyên ngành Công trình thủy để công tác dịch thuật đạt độ chuẩn xác cao.

Giảng viên dạy ngoại ngữ cũng có thể sử dụng danh mục này làm tư liệu biên soạn các sách tiếng Anh thực hành kết ngữ chuyên ngành. Người biên soạn sách nên chia nhỏ danh mục theo từng chủ đề môn học để tạo điều kiện thuận lợi hỗ trợ cho giảng viên dạy chuyên ngành kiểm tra kiến thức của sinh viên về từ chuyên ngành trong mỗi học phần giảng dạy.

Về mặt phương pháp, các bước trong quy trình chiết xuất và kiểm định kết ngữ trong đề án thử nghiệm này hoàn toàn có thể được tiến hành tương tự cho các dự án mở rộng cho toàn ngành Công trình, ngành Hàng hải, hoặc bất kỳ chuyên ngành nào khác.

TÀI LIỆU THAM KHẢO

- [1] Martyńska, Małgorzata. "Do English language learners know collocations?". *Investigationes linguisticae*. Vol. 11. pp. 1-12, 2004.
- [2] Nesselhauf, Nadja. *Collocations in a learner corpus*. Amsterdam: John Benjamins, 2005.
- [3] Bahns, Jens, and Moira Eldaw. "Should we teach EFL students collocations?". *System* 21. Vol. 1. pp. 101-114, 1993.
- [4] Hill, Jimmie. "Revising priorities: From grammatical failure to collocational success". *Teaching collocation*. pp. 47-69, 2000.
- [5] Aisenstadt, Ester. "Collocability restrictions in dictionaries". *ITL-International Journal of Applied Linguistics*. Vol. 45(1), pp. 71-74, 1979.
- [6] Hunston, Susan. *Corpora in applied linguistics*. Ernst Klett Sprachen, 2002.
- [7] Kilgarriff, Adam. *Sketch Engine*. <http://www.sketchengine.eu>, 2014.
- [8] Scott, Mike. *WordSmith Tools Version 7.0*. 0.54. 2016.
- [9] UCREL. *Free CLAWS web tagger*. ucrel-api.lancaster.ac.uk/claws/free.html, 2014.
- [10] ANC. *Open American National Corpus*. <http://www.anc.org/data/oanc/download/>, 2015.
- [11] Fang, Zhihui, Mary J. Schleppegrell, and Beverly E. Cox. "Understanding the language demands of schooling: Nouns in academic registers". *Journal of Literacy Research*. Vol. 38 (3), pp. 247-273, 2006.
- [12] Gulec, Nuray, and Bulent Arif Gulec. "Lexical collocations (verb+ noun) across written academic genres in English". *Procedia-Social and Behavioral Sciences*. Vol. 182, pp. 433-440, 2015.
- [13] Ackermann, Kirsten, and Yu-Hua Chen. "Developing the Academic Collocation List (ACL)-A corpus-driven and expert-judged approach". *Journal of English for Academic Purposes*. Vol. 12 (4), pp. 235-247, 2013.

Ngày nhận bài: 25/12/2019
 Ngày nhận bản sửa: 02/01/2020
 Ngày duyệt đăng: 09/01/2020