Dual-Tier Sampling Methodology for Extracting and Analysing Metaphorical Terminology: Evidence from Welsh Educational Terms *

Catrin Owen[1,*,†] and Sarah Cooper[2,†]

[1] *Bangor University, Language Technologies Unit, Bangor University, College Road, Bangor, Wales*
[2] *Bangor University, School of Arts, Culture and Language, Bangor University, Bangor, Wales*

Abstract

This paper examines how contrasting two sampling methodologies reveals different patterns when analysing metaphorical terminology in multidisciplinary terminological resources. Using Welsh educational terminology from *Y Termiadur Addysg* we compare two complementary approaches: (1) semantic domain-based sampling (Tier 1: 160 terms from common everyday source domains—HUMAN BODY, ANIMALS, CLOTHING, DOMESTIC ENVIRONMENT, FOOD), and (2) subject-specific sampling (Tier 2: 166 terms from Geography/Geology, ICT/Computing, Sociology).

The two sampling approaches reveal systematically different structural and semantic patterns. Semantic domain-based sampling predominantly captures noun+noun compounds and appearance-to-appearance mappings. Subject-specific sampling reveals greater structural and semantic diversity, including verbal metaphors (entirely absent from Tier 1), increased single-noun forms and greater concept-to-concept (i.e. abstract) mappings. Comparison between approaches indicates that each sampling strategy foregrounds different aspects of metaphorical terminology, suggesting that some of the observed patterns may be influenced by the sampling design.

These findings highlight that observed patterns in metaphorical terminology are shaped by sampling strategy, with each approach providing context for interpreting the other. Combining domain-based and subject-specific sampling enables researchers to distinguish sampling artifacts from genuine terminological patterns, with implications for termbase design and standardisation practices in multilingual contexts.

Keywords

sampling methodology, metaphorical terminology, Welsh terminology, multilingual terminology, educational terminology

1. Background: metaphor in terminology

Metaphor constitutes a fundamental cognitive mechanism that allows people to understand unfamiliar or abstract concepts through systematic mappings from familiar, concrete domains [1]. In terminology research, metaphor shapes the way technical and educational concepts are named, influencing how these terms are stored, accessed and interpreted in terminological databases.

Investigating these metaphorical patterns in terminology requires selecting relevant terms from databases. Previous research on metaphorical terminology employs methodologically diverse sampling approaches, including manual dictionary-based searches [2, 3, 4, 5], restricted alphabetical sampling [6], corpus-based methods [7], and semi-automatic retrieval from annotated corpora [8]. These approaches therefore differ substantially in data source (dictionaries, databases, corpora), sampling logic (alphabetical, keyword-based, domain-based, field-specific), and level of automation, yet most studies employ single sampling strategies.

Despite growing research on metaphor in specialised terminology [9, 10, 11, 12, 13], systematic methodological comparison examining how different sampling approaches reveal different metaphorical characteristics remains underexplored. Different methodologies may systematically

* *5th International Conference on "Multilingual digital terminology today. Design, representation formats and management systems" (MDTT) 2026, June 25-26, 2026, Zadar, Croatia.*
[1*] Corresponding author.
[†] These authors contributed equally.
c.owen@bangor.ac.uk (C. Owen); s.cooper@bangor.ac.uk (S. Cooper)

affect the detection and analysis of metaphorical patterns. This gap is particularly important for multilingual terminology management, where understanding the full range of metaphorical patterns, not just those visible through single methodology approaches, informs our knowledge of standardisation practice and termbase architecture requirements.

By comparing domain-based[1] and subject-specific approaches within a single terminological resource, this study examines how methodological choices influence both the types of metaphors observed and their potential representation in terminological resources. Using Welsh educational terms as a case study, we examine what distinct structural and semantic patterns emerge from each sampling approach, how contrasting findings from both approaches reveals what each systematically captures or excludes, and what implications these findings have for metaphorical terminology analysis, terminology extraction practice and termbase design and standardisation.

2. Methodology

2.1. Data

We extracted 326 metaphorical terms from *Y Termiadur Addysg* (TA) [14], the authoritative Welsh-English educational terminology database maintained by the Language Technologies Unit, Bangor University and funded by the Welsh Government. The database contains approximately 56,000 standardised entries across curriculum subject areas, providing a comprehensive resource for examining metaphorical terminology in the Welsh educational context.

2.2. Sampling methodology

Two sampling strategies were employed to extract the metaphorical terminology. The first, domain-based sampling (Tier 1), drew 160 metaphorical terms by manually searching the termbase for lexical items from five semantic source domains: HUMAN BODY (*pen* 'head', *braich* 'arm', *troed* 'foot', etc.), ANIMALS (*ci* 'dog', *broga* 'frog', *llygoden* 'mouse', etc.), CLOTHING (*siaced* 'jacket', *ffedog* 'apron', *sgarff* 'scarf', etc.), DOMESTIC ENVIRONMENT (*drws* 'door', *ffenestr* 'window', *wal* 'wall', etc.), and FOOD (*brechdan* 'sandwich', *reis* 'rice', *afal* 'apple', etc.)

These domains are documented as highly productive metaphor sources cross-linguistically [6, 3, 15, 16] and are grounded in universal and embodied experience. Sampling by searching and examining items from these established source domains therefore ensures sufficient metaphorical term retrieval, while also capturing systematic patterns in how specific semantic domains contribute to terminological metaphor. However, given the noun-centric nature of these entity-focused domains (body parts, animals, objects), it was predicted that this sampling strategy would predominantly yield nominal metaphorical structures.

The second sampling strategy, subject-specific sampling (Tier 2), adopts a complementary approach. This strategy extracted 166 metaphorical terms through manual exhaustive examination of three subject areas: Geography and Geology (physical and descriptive, addressing observable phenomena), ICT and Computing (process-oriented dealing with abstract digital operations), and Sociology (abstract, theoretical field addressing social structures and relationships). These subject areas were selected to investigate whether different subject characteristics (concrete vs. abstract, entity-based vs. process-based) reveal subject-specific metaphorical patterns.

2.3. Metaphor identification and analysis

The metaphoricity of sampled terms was confirmed using the Metaphor Identification Procedure VU (MIPVU) [16]. All confirmed metaphorical terms across Tier 1 (Domain-based) and Tier 2 (Subject-specific) were then systematically analysed according to their structural and semantic features following a multi-dimensional coding framework primarily adapted from (2022, 474) [3].

[1] Throughout this paper, following the use of the term "source domain" in Conceptual Metaphor Theory, *domain* refers to semantic domains (conceptual categories such as HUMAN BODY, ANIMALS) rather than subject domains or academic disciplines.

The structural analysis considered the parts of speech of the constituent elements of each term (e.g. noun+noun, noun+adjective, verbal noun). The semantic analysis examined the type of metaphorical mapping present, such as appearance-to-appearance (based on visual similarity, e,g, *lafa clustog* – 'pillow lava'), concept-to-concept (based on abstract similarities, e.g. *priodas gragen wag* – 'empty shell marriage'), and function-to-function (based on functional similarities, e.g. *siaced lwch* – 'dust jacket').

The analysis presented here focuses on the broader structural and semantic patterns that emerge within each sampling tier. The results below are presented as comparative overviews, focusing on tendencies and contrasts between Tier 1 (Domain-based) and Tier 2 (Subject-specific). This approach reveals how different sampling strategies shape the types of metaphorical patterns observed.

3. Results

3.1. Structural patterns: parts of speech

As shown in Table 1, domain-based sampling strongly favoured noun + noun compounds (40.88%, n=65), exemplified by *pwyth cadwyn* 'chain stitch', *graff olwyn* 'pie chart', *ysgol DNA* 'DNA ladder', and *gwregys asteroid* 'asteroid belt', where concrete objects provide descriptive specification for technical concepts. Single noun forms comprised 19.49% (n=31), including *llygoden* 'mouse' (computing) and *mwydyn* 'worm' (malware).

Subject-specific sampling revealed a different distribution. Single-noun forms increased substantially (26.5%, n=44), with complete conceptual transfers such as *basn* 'basin', *firws* 'virus', *cwmwl* 'cloud', and *cwlwm* (lit. knot) 'bond'. Noun+noun compounds appeared with reduced frequency (21.08%, n=35). Critically, verbal noun forms (i.e. non-finite verb forms that also functions nominally) emerged in Tier 2 (7%, n=11)—*pori* 'to browse/graze', *llwytho* 'to load', *cipio* 'to capture'. Such verbal noun forms were entirely absent from domain-based sampling.

Table 1
Structural pattern comparison between sampling tiers (selected patterns)[2]

Pattern	Tier 1 (Domain-based)	%	Tier 2 (Subject-specific)	%
Noun+noun	65	40.88	35	21.08
Single noun	31	19.49	44	26.5
Noun + adjective	11	6.9	31	18.67
Verbal noun	0	0	11	7.0

The contrast reveals sampling-dependent patterns in both directions. The entity-focus of domain-based sampling (BODY, ANIMALS, OBJECTS) naturally yields nominal structures, capturing metaphors where concrete objects specify technical concepts. However, in the subject-specific approach there is an increased frequency of single-noun forms (26.5%) and the presence of verbal metaphors (7%). These become interpretable as significant patterns only through contrast with their distribution in Tier 1. The complete absence of verbal forms in Tier 1 (0%) suggests this is a systematic sampling constraint. Fields like Computing and Sociology require describing operations and processes (e.g., *pori* 'to browse', *llwytho* 'to load' in Computing; *sianelu* 'canalization' in Sociology). Verbal noun forms meet these conceptual needs, but they are not captured by entity-

[2] Table shows the four most prominent structural patterns. Other patterns (including noun+noun+adjective, verbal noun + noun, and compound forms) account for the remaining 32.73% (Tier 1) and 26.75% (Tier 2).

focused domain sampling using noun-centric semantic fields. Overall, the comparison of the structural results for Tier 1 (domain-based) and Tier 2 (subject-specific) sampling suggests that each approach highlights different structural properties of metaphorical terminology. Domain-based sampling tends to emphasise noun+noun compounds, while subject-specific sampling reveals greater structural diversity.

3.2. Semantic patterns: mapping types

The domain-based sampling in Tier 1 strongly revealed appearance-to-appearance mappings (44.37%, n=71), where visual resemblance motivates metaphor: *ysgol DNA* 'DNA ladder' (double helix structure resembles ladder rungs), *graff olwyn* 'pie chart' (lit. 'wheel graph' - circular segmented diagram resembling a wheel), *llwyfan ffedog* 'apron stage', and *rhybed pen madarch* 'mushroom rivet'. Concept-to-concept (i.e. abstract) mappings appeared substantially less frequent (11.9%, n=19), including *priodas gragen wag* 'empty shell marriage' and *ffenestr drosglwyddo* 'transfer window'. Position-to-position mappings (8.8%, n=14) appeared in positional metaphors like *troednodyn* 'footnote' and *gwadn rhewlif* 'glacier sole', drawing from HUMAN BODY and ANIMALS domains where body-part position provides salient features. Movement-to-movement mappings concentrated heavily in Physical Education contexts (with 88.8% of these movement mappings drawn from the ANIMALS domain): *naid broga* 'frog jump', *llam llyffant* 'leapfrog', *nofio ci* 'dog paddle'. Combined mappings (multiple simultaneous correspondences) appeared in 22.6% (n=36), most frequently appearance+function (n=15), as in *crafanc* 'chuck' (lit. claw – clamping device in machinery), which both resembles and functions like a claw.

The subject-specific sampling in Tier 2 revealed markedly different semantic profiles. While Geography/Geology maintained appearance-to-appearance patterns consistent with describing physical landforms (*pen mynydd* 'mountain head', *nodwydd lafa* 'lava needle'), ICT/Computing and Sociology showed concept-to-concept mappings as predominant. ICT examples include *traffig data* 'data traffic' (information flow as vehicular movement), *ymosodiad nerth pur* 'brute force attack' (computational trial-and-error as physical bombardment), and *porth* 'port' (data entry/exit point). Sociology showed the highest concentration of concept-to-concept mappings: *haenu cymdeithasol* 'social stratification' (social hierarchy as physical layering), *peirianneg gymdeithasol* 'social engineering' (social change as mechanical process), and *damcaniaeth bath cynnes* 'warm bath theory' (home life as comfort from warm water). Combined mappings in Tier 2 showed different patterns than Tier 1, with concept+visual-spatial (*crogfur* 'hanging wall', *blaendir* 'foreground') and concept+action (*haciwr* 'hacker') emerging prominently.

The contrast reveals how each sampling strategy foregrounds fundamentally different semantic patterns. The domain-based sampling favoured appearance-mapping (44.37%), reflecting the perceptual salience of entity-focused domains (BODY, ANIMALS, and OBJECTS) which offer visible features that ground appearance-based correspondences. Without subject-specific sampling, this pattern might suggest that Welsh educational terminology generally favours appearance-based metaphor. However, subject-specific sampling highlights different patterns suggesting this appearance-based preference may be due to the difference in sampling strategy. The abstract target domains in Sociology cannot be metaphorised through appearance because social structures lack perceptual form, requiring concept-to-concept mappings, while process-oriented concepts in Computing require mappings of operations and flows rather than visual resemblances. Conversely, the domain-based sampling from Tier 1 reveals patterns that subject-specific sampling might miss, such as the concentration of movement-to-movement mappings from ANIMALS in Physical Education (88.8%). This demonstrates systematic source domain specialisation that subject-specific sampling does not capture. Each sampling approach thus illuminates the other's value: domain-based sampling quantifies source domain productivity across fields, while subject-specific sampling reveals how target domain epistemology shapes mapping requirements.

4. Discussion

This study sets out to examine how different sampling strategies within a Welsh educational terminology resource shape the metaphorical patterns that become visible for analysis. By comparing domain-based and subject-based specific sampling of Welsh educational terminology, we demonstrate that metaphor analysis is sensitive to the method of term selection within the resource. The sampling strategy has consequences for the structural and semantic profile of metaphorical terminology, which raises questions about how metaphorical terminology is developed, interpreted and standardised.

The structural differences observed between the two sampling techniques highlight how terminological organisation shapes which metaphorical forms are foregrounded. Domain-based sampling favours complex nominal structures and entity-based metaphors, aligning closely with prioritising nouns for concept representation and standardisation. This may explain why verbal and process-oriented metaphors are underrepresented in studies that choose to rely on predefined source domain lists: they are predisposed to capture nominal constructs because of the nature of the source domains. From a terminology perspective, this raises questions about the representational completeness, as process metaphors play a crucial role in fields where procedural knowledge is central, which was highlighted with the subject-based sampling in this study.

Differences between the domain-based and subject-based sampling further indicate that the method of sampling foregrounds different cognitive functions of metaphor in the terminology. The domain-based sampling privileged visually grounded and appearance-based metaphors, which highlight anchoring new concepts in familiar, concrete domains. In contrast, the subject-specific sampling revealed a wider range of metaphorical strategies dependent on the disciplinary norms, including abstract conceptual mappings. From a terminological standpoint, relying on one search approach may favour certain strategies over others, potentially shaping how concepts are presented to users and how effectively terminology supports understanding across different subject areas.

Taken together, researchers analysing metaphorical terminology in dictionaries and termbases should be aware of the value of combining sampling strategies where resources permit. Domain-based and subject-specific approaches yield systematically different representations of metaphorical terminology, affecting which forms, mappings, and conceptual relations are documented. A combined approach offers a more balanced perspective, particularly when termbases provide subject-field filtering alongside lexical search capabilities. This has practical implications for termbase design, standardisation practices, and content representativeness, particularly in multilingual and specialised contexts where the visibility of certain metaphorical patterns may directly influence how concepts are understood. For termbase architectures, implementing subject field filters or domain tags would enable researchers to conduct both sampling approaches systematically, supporting more comprehensive metaphor analysis within a single resource.

The study is necessarily limited in scope, focusing on a single terminological resource and a restricted set of subject areas within Welsh educational terminology. Semantic categorisation was intentionally simplified to remain within the constraints the conference aims. Future research could expand subject coverage, adopt more fine-grained mapping analyses, and explore cross-linguistic comparisons with English or other languages to examine how metaphorical strategies are realised in multilingual terminological infrastructures.

5. Conclusions

This study comparing domain-based and subject-specific sampling within a Welsh educational terminological resource offers a useful methodological perspective on metaphorical terminology. The analysis of Welsh educational terms indicates that each approach foregrounds different structural and semantic tendencies: domain-based sampling highlights entity-based, appearance-

oriented metaphors and systematic source-domain productivity, while subject-specific sampling captures a broader range of forms, including verbal metaphors and field-specific strategies. Examining both sampling techniques together highlights how some recurring patterns may reflect sampling method rather than general properties of terminology itself. For researchers working on metaphor in terminological resources, this suggests that combining complementary sampling strategies can support more cautious interpretation by distinguishing sampling-dependent patterns from terminology-dependent ones. For terminology development and management, the findings highlight the importance of accommodating diverse metaphorical strategies shaped by disciplinary norms, particularly in multilingual and multidisciplinary termbases.

Declaration on Generative AI

During the preparation of this work, the authors used Claude Sonnet 4.5 (Anthropic, 2025) for: text translation, grammar and spelling check, paraphrase and reword. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] G. Lakoff, M. Johnson, Metaphors We Live By, University of Chicago Press, Chicago, 1980.

[2] M. Siqueira, A. F. Souto de Oliveira, D. Dienstbach Hubert, G. Faé de Almeida, L. M. Brangel, Metaphor identification in a terminological dictionary, Ibérica 17 (2009) 157–174.

[3] Y. Hrybinyk, T. Halai, N. Yesypenko, I. Bloshchynskyi, Approaching metaphorical terms in subject-specific terminologies (geologic and geodetic): Semantic and structural aspects, World Journal of English Language 12 (2022) 470-484. doi:10.5430/wjel.v12n6p470.

[4] E. Khabirova, Metaphorical terms in the context of linguistic research articles, XLinguae 11 (2018) 499–507. doi:10.18355/XL.2018.11.02.40.

[5] A. Roldán-Riejos, G. Cuadrado, Metaphor and figurative meaning construction in science and technology (English and Spanish), Procedia – Social and Behavioral Sciences 212 (2015) 271–277. doi:10.1016/j.sbspro.2015.11.348

[6] J. Tretjakova, The Role of Metaphor in Comprehension of Railway Terminology, in: L. Ilynska, M. Platanova (Eds.), Meaning in Translation: Illusion of Precision, Cambridge Scholars Publishing, Newcastle upon Tyne, 2016, pp. 197–210.

[7] M. Tercedor Sánchez, J. M. Ureña Gómez-Moreno, J. A. Prieto Velasco, Grasping metaphoric and metonymic processes in terminology, JoSTrans: The Journal of Specialised Translation 18 (2012) 187-205. doi:10.26034/cm.jostrans.2012.445.

[8] J. M. Ureña Gómez-Moreno, P. Faber, Strategies for the semi-automatic retrieval of metaphorical terms, Metaphor and Symbol 26 (2011) 23–52. doi:10.1080/10926488.2011.535415.

[9] P. Faber, C. I. Lopez-Rodriguez, Terminology and specialised language, in: P. Faber (Ed.), A Cognitive Linguistics View of Terminology and Specialized Language, De Gruyter Mouton, Berlin, 2012, pp. 9–32.

[10] B. Alexiev, Towards an experientialist model of terminological metaphorisation, Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 10 (2004) 189–213. doi:10.1075/term.10.2.03ale.

[11] R. Temmerman, Towards New Ways of Terminology Description: The sociocognitive approach, John Benjamins Publishing Company, Amsterdam/Philadelphia, 2000. doi:10.1075/tlrp.3.

[12] J. M. Ureña Gómez-Moreno, F. J. Ruiz de Mendoza Ibáñez, Conceptual metaphors in terminology and specialised language discourse, in: P. Faber, M.-C. L'Homme (Eds.), Theoretical Perspectives on Terminology. Explaining terms, concepts and specialized knowledge, John Benjamins Publishing Company, Amsterdam/Philadelphia, 2022, pp. 377–396. doi:10.1075/tlrp.23.17gom.

[13] Y Termiadur Addysg, Welsh terminology dictionary for education, 2024. URL: https://www.termiaduraddysg.cymru/.
[14] B. A. Barkhramovich, Structural and semantic analysis of English metaphors in transport terminology, International Journal of Trend in Scientific Research and Development (IJTSRD) 6 (2022) 1716–1717. Unique Paper ID: IJTSRD49833.
[15] I. Molchanova, V. Kurilenko, Y. Biriyukova, E. Kulikova, M. Makarova, Specifics of translating metaphors from English Into Russian in medical discourse, in: F. Uslu (Ed.), INTCESS 2019 – 6th International Conference on Education and Social Sciences, Ocerint Publishing, Istanbul, 2019, pp. 122–128.
[16] G. J. Steen, A. G. Dorst, J. B. Herrmann, A. A. Kaal, T. Krennmayr, T. Pasma, A Method for Linguistic Metaphor Identification: From MIP to MIPVU, John Benjamins Publishing Company, Amsterdam/Philadelphia, 2010, doi:10.1075/celcr.14.