International Journal of English Literature and Social Sciences
Vol-10, Issue-4; Jul-Aug, 2025

Peer-Reviewed Journal
Journal Home Page Available: https://ijels.com/
Journal DOI: 10.22161/ijels

Integrating Conceptual Metaphor Theory and Corpus-Based Data: A Stepwise Application of the Single-Lexeme Model of Analysis

Mehdi Morchid

PhD in Cognitive Linguistics, *Ibn Tofail University*, Kenitra, Morocco

Received: 22 Jun 2025; Received in revised form: 16 Jul 2025; Accepted: 20 Jul 2025; Available online: 28 Jul 2025

Abstract— *This paper presents a stepwise methodological framework for integrating Conceptual Metaphor Theory (CMT) with corpus-based data using a single-lexeme model of analysis. Building on Stefanowitsch's Metaphorical Pattern Analysis (MPA), the approach begins with the selection of a target domain and a single denotative lexeme, from which metaphorical patterns are retrieved and interpreted within concordance lines. The process combines the empirical strengths of corpus linguistics with the interpretive depth of CMT by identifying recurring cross-domain mappings and subsuming them under conceptual metaphors. To address challenges of subjectivity in metaphor identification, the Pragglejaz Metaphor Identification Procedure is integrated into the analysis. Additionally, the digital platform Atlas.ti is employed to manage, code, visualise and synthesise metaphorical data, offering a robust environment for both qualitative and quantitative exploration. This methodological integration supports greater transparency, reproducibility and interpretive precision in metaphor research. The paper concludes by reflecting on epistemological tensions between data-driven linguistics and cognitively oriented metaphor theory, calling for careful justification of analytic decisions within such interdisciplinary research.*

Keywords— ***Conceptual Metaphor Theory, Corpus Linguistics, Metaphorical Pattern Analysis, Single-Lexeme Model, Pragglejaz Procedure, Cognitive Linguistics, Qualitative Data Analysis.***

I. INTRODUCTION

There is a strong methodological compatibility between corpus linguistics and the analytic demands of Conceptual Metaphor Theory (CMT). In Words and Their Metaphors: A Corpus-Based Approach, Stefanowitsch (2006) outlines both theoretical and practical continuities between metaphor analysis and corpus-based data. Theoretically, both approaches reject the notion of language as a neutral conduit for communication. Instead, they view language as a reflection of underlying cognitive structures accessible through analysis. CMT holds that metaphors function as cognitive mechanisms for structuring abstract target domains in terms of more concrete source domains. Similarly, corpus linguistics focuses on the detection of recurring patterns in language, which are construed as evidence of an interconnection between linguistic and cognitive components. Practically, corpus methods enable the study of phenomena such as entrenchment through an empirical assessment of metaphor frequency and distribution, bringing the stabilising elements of reproducibility and standardisation to the inherently interpretive nature of conceptual metaphor analysis.

One of the key strengths of corpus-based metaphor research is its ability to capture contextual variation. According to Stefanowitsch (2006), the use of metaphor is influenced by discourse type, authorial stance and communicative purpose. These contextual dimensions become more visible in large-scale corpora than in artificially constructed datasets such as interviews or surveys. Moreover, because corpus data are typically

created independently of the research process, they are less susceptible to manipulation or researcher bias.

Stefanowitsch (2006) also underscores the role of corpus linguistics in integrating quantitative and qualitative approaches to metaphor analyses. It allows for statistical studies of frequency and distribution while supporting interpretive inquiry into the meanings, ideologies and evaluative functions of metaphors. The interplay of these approaches makes corpus-based metaphor studies particularly effective at uncovering how metaphors contribute to broader discourse functions.

II. STEFANOWITSCH'S METAPHOR PATTERN ANALYSIS

Stefanowitsch (2006) offers a detailed model for analysing conceptual metaphors in large corpora. His method, known as Metaphorical Pattern Analysis (MPA), systematically investigates cross-domain mappings in corpus data. MPA begins with a target-domain-focused approach where the researcher selects a target domain, identifies a set of lexemes associated with it and retrieves their occurrences in the corpus. This brings more focus than bottom-up metaphor analyses where the researcher proceeds through the text linearly identifying metaphors as they appear. Hence, instead of collecting all metaphorical expressions in a text, the MPA tracks target-domain lexemes in context and examines the conceptual metaphors under which they are subsumed.

The process involves selecting a target domain, identifying relevant lexemes that evoke it and extracting concordance lines. The analysis then focuses on the source domains used to structure the target domain and identifies the specific elements of meaning that are mapped onto it. Metaphors involving adjacent or similar source domains are grouped together and synthesised into conceptual metaphors following Lakoff and Johnson's (1980) format, such as ARGUMENT IS WAR (Lakoff & Johnson, 1980). MPA incorporates quantitative elements like measuring metaphor frequency to determine degrees of entrenchment and qualitative interpretive dimensions such as the reconstruction of how theories are built through conceptual metaphors. The method detects which specific mappings are most often reproduced. Additionally, the focus on multiple lexemes denoting the same target domain allows for comparisons across languages, registers and disciplines.

Compared to introspective methods, MPA enables the retrieval of a broader range of metaphors and provides analytical focus through the designation of a single target domain. It facilitates the empirical analysis of abstract concepts by anchoring them in measurable lexemes and corpus data. This possibility of integrating quantitative frequency analysis with qualitative interpretation yields both precision and conceptual depth.

Despite its logical coherence and intuitive applicability, the method has limitations. It overlooks metaphors that do not explicitly reference the target domain lexically. Additionally, even with a thorough selection of target domain lexemes, there remains the risk of missing relevant terms, especially if obscure or context-dependent words are not pre-selected as indicators of the target domain. Consequently, some metaphors may go undetected if they rely on unexpected lexical items.

III. A SINGLE LEXEME SPECIFIC APPLICATION OF STEFANOWITSCH'S MPA

The single lexeme approach represents a specific application of Stefanowitsch's broader method (MPA). This focused implementation begins with the designation of a target domain. However, instead of tracking multiple lexemes associated with that domain, it limits the analysis to a single word. This word may either eponymously refer to the target domain or be central to its meaning. All occurrences of the selected lexeme are retrieved from the corpus within their concordance lines. Emphasis is placed on identifying recurrent metaphorical constructions involving various source domains. These constructions, which Stefanowitsch (2006) refers to as "metaphorical patterns," constitute the core of the analysis. Subsequently, these patterns are subsumed under broader conceptual metaphors, which are conventionally written in capital letters following the tradition established by Lakoff and Johnson (1980).

This phase of the analysis is interpretive in nature, as conceptual metaphors are not directly retrieved from the corpus but are rather formulated through the researcher's interpretive choices. Nonetheless, the process remains data-driven, since it is grounded in recurring patterns observed in the corpus data. The resulting conceptual metaphors can be used in various ways, depending on the research's aim (Semino, 2008). For example, they may reveal how authors construct worldviews, ideological positions or evaluative stances. Alternatively, the analysis may serve as the basis for comparing competing metaphorical framings of the same target domain, tracing diachronic changes or conducting multimodal analyses that connect linguistic metaphors with visual or symbolic elements.

Within Stefanowitsch's model, metaphoricity is determined manually on a case-by-case basis. Nearly two decades after his article, there is still no fully-automated

tool that can reliably identify metaphoricity. Shutova (2015) surveys a number of digital platforms that assist in metaphor research, but highlights their limited reliability in distinguishing literal from metaphorical usage. Manual analysis therefore remains indispensable, though it introduces subjectivity and poses challenges to reproducibility. To address this, researchers in cognitive linguistics have developed standardised methods for metaphor identification. One such method is the Pragglejaz Metaphor Identification Procedure (Group Pragglejaz, 2007), which reduces interpretive bias by providing explicit, replicable steps for determining metaphoricity.

IV. USING THE PRAGGLEJAZ METAPHOR IDENTIFICATION PROCEDURE (MIP)

A persistent challenge in metaphor research lies in determining whether an expression qualifies as metaphorical. Without a clear method, this process is vulnerable to the researcher's subjective judgment, which compromises academic rigour. Subjectivity undermines reliability as decisions regarding metaphoricity may vary across researchers, thereby weakening the reproducibility, triangulation and verification of findings. When metaphors are identified based on intuition rather than systematic procedures, the resulting analysis risks inconsistency and limits the study's contribution to cumulative knowledge. In contrast, objectivity ensures that research outcomes are not shaped by personal bias, but by replicable methods. It enables different researchers, applying the same procedures to the same data, to arrive at similar conclusions. This is particularly important for peer review, which evaluates the logical coherence of a study rather than the researcher's opinions.

To address this methodological concern, the Pragglejaz Metaphor Identification Procedure (Group Pragglejaz, 2007) offers a systematic framework for identifying metaphorical language. Developed by a team of scholars, the procedure aims to reduce subjectivity by providing clear steps to determine whether a lexical unit is used metaphorically. Importantly, the Pragglejaz Metaphor Identification Procedure does not attempt to infer the speaker's intention or the effect on the listener. It focuses on the linguistic usage itself. Drawing on the foundational work of Lakoff and Johnson (1980), the procedure is grounded in the premise that linguistic metaphors are surface manifestations of deeper conceptual metaphors.

The Pragglejaz Metaphor Identification Procedure consists of five main steps designed to systematically determine whether a lexical unit is used metaphorically (Group Pragglejaz, 2007).

1. Read the entire text to develop a general understanding of its meaning.
2. Identify the targeted lexical units in the text. These are typically single words or short fixed expressions.
3. Determine the contextual meaning of each lexical unit, describing how it is used in the specific context of the text. This includes considering the immediate linguistic co-text (the words before and after it).
4. Determine whether the lexical unit has a more basic meaning in other contexts. A "basic" meaning is typically more concrete, physical or related to bodily action, and it is historically older.
5. Compare the contextual and basic meanings. If the contextual meaning contrasts with the basic meaning, but can be understood in comparison with it, then the lexical unit is metaphorical. If there is no such contrast, or if the contextual meaning is simply an extension of the basic meaning, then the word is not considered metaphorical.

The Pragglejaz Metaphor Identification Procedure enhances objectivity by replacing intuition with a replicable decision-making process grounded in linguistic evidence. Although it cannot offer perfectly definitive results, since it remains a tool of the humanities, it significantly reduces subjectivity in metaphor identification. One of the procedure's key strengths is its adaptability to both qualitative and quantitative research contexts. Some approaches aim to combine the researcher's sensitivity with computational tools, such as the MIPVU, Metaphor Identification Procedure Vrije Universiteit (Steen, Dorst, Herrmann, Kaal & Krennmayr, 2010), which integrates detailed annotation protocols with automated or machine-assisted methods. However, fully automated metaphor detection remains imperfect, and the Pragglejaz procedure continues to offer value through its rigorous, researcher-led methodology.

V. USING THE DIGITAL PLATFORM ATLAS.TI TO ASSIST IN THE ANALYSIS OF CONCEPTUAL METAPHORS

Corpus linguistics has seen the emergence of various digital platforms designed to assist in the analysis of conceptual metaphors. These platforms offer varying levels of visualisation, automation and text mining. The following is a detailed account of how Atlas.ti could be employed in conceptual metaphor analysis. The decision to use Atlas.ti is based on its capacity to efficiently handle large quantitative and qualitative datasets as well as its robust coding tools. While the majority of text-processing programmes that are commonly used in corpus linguistics are designed to prioritise quantitative, numerical output,

Atlas.ti offers the researcher the ability to carry out nuanced interpretive work. This makes the platform particularly well-suited to the analysis of conceptual metaphors and the integration of the Pragglejaz Metaphor Identification Procedure (Group Pragglejaz, 2007) to determine metaphoricity.

The Corpus data are imported into Atlas.ti as Primary Documents (PDs), which are documents made ready for segmentation, reorganisation and retrieval within the software. Using the "Search and Query" functions, all instances of the designated token are identified. The researcher, then, applies the Pragglejaz Metaphor Identification Procedure to isolate only metaphorical uses of the token (Group Pragglejaz, 2007). The "Quotation" feature enables the extraction of all concordance lines containing metaphors involving the token. Using the "Coding" feature, each metaphor is analysed by identifying its source and target domains and reconstructing its cross-domain mappings. Metaphors that are shaped by similar source domains are grouped as "metaphor patterns" using the "Hermeneutic Units" feature (Stefanowitsch, 2006). The "Co-occurrence Models" tool helps reveal relationships between codes and the contexts in which they appear. Finally, the "Network View" feature visually displays all quotations, codes and hermeneutic units, making it easier to identify connections among conceptual metaphors and the themes to which they contribute. This variety of features makes Atlas.ti effective in automating laborious processes.

Konopásek (2007) identifies four guiding principles in the design of Atlas.ti: "exploration, visualisation, integration, and serendipity" (2007). Exploration refers to the researcher's capacity to intervene in the data using tools for segmentation, coding and annotation, allowing the uncovering of deeper layers of meaning. Visualisation concerns the representation of abstract relationships using tools such as Network Views, Co-occurrence Models and Word Clouds. This visual capacity facilitates the detection of patterns and connections within the data. Integration allows for the unification of disparate data elements into meaningful categories, as exemplified by Hermeneutic Units, which can encapsulate everything from direct quotations to interpretive annotations. As new meanings emerge, the integration process supports continual refinement and the preservation of analytic nuance. Serendipity, as described by Friese represents the highest order of analysis (2019). It enables the discovery of latent patterns or relationships not previously considered. Atlas.ti is designed to foster such insights by revealing connections and structures that might otherwise remain obscured. This feature is especially valuable in metaphor analysis, as conceptual metaphors inherently rely on implicit cross-domain mappings. Uncovering these hidden links often leads to unexpected, yet theoretically significant findings (Lakoff & Johnson, 1980).

VI. ENSURING TRUSTWORTHINESS

To ensure trustworthiness, several strategies could be employed. These include involving multiple coders who use the same analytical framework and continuously comparing emergent themes with existing scholarship. The integration of Computer-Assisted Qualitative Data Analysis Software (CAQDAS), specifically Atlas.ti, streamlines the mechanical aspects of data management and allows greater focus on interpretive analysis. The methodological framework detailed in this paper provides a robust and coherent structure for investigating how conceptual metaphors operate in corpus-data.

VII. CONCLUSION

There is both theoretical and practical alignment between corpus-based methodologies and CMT. This paper has focused on a specific application of Stefanowitsch's MPA. The single-lexeme model brings more focus by letting the researcher select a single word, locate its metaphorical uses throughout the corpus and subject it to analysis. The process incorporates the empirical strength of corpus-based data and the depth of interpretive analysis characteristic of CMT.

The employment of the Pragglejaz Metaphor Identification Procedure (2007) mitigates subjectivity in determining metaphoricity. Despite the necessity of hands-on involvement of the researcher in metaphor analysis, the digital platform Atlas.ti offers substantial options to facilitate data management, visualisation and retrieval.

The framework detailed in this paper provides a coherent and efficient modus operandi for a methodologically rigorous and theoretically informed analysis of conceptual metaphors. Nonetheless, challenges persist. While corpus linguistics operates within a paradigm of quantifiable data, CMT engages abstract cognitive constructs. This epistemic divergence necessitates detailed justification for every methodological decision the researcher makes.

REFERENCES

[1] Friese, S. (2019). *Qualitative data analysis with ATLAS*. ti.

[2] Group, P. (2007). *MIP: A method for identifying metaphorically used words in discourse*. Metaphor and symbol, 22(1), 1-39.

[3] Konopásek, Z. (2007). *Making thinking visible with Atlas*. ti: Computer assisted qualitative analysis as textual practices. Historical Social Research/Historische Sozialforschung. Supplement, 276-298.

[4] Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.

[5] Semino, E. (2008). *Metaphor in discourse (p*. 81). Cambridge: Cambridge University Press.

[6] Shutova, E. (2015). Design and evaluation of metaphor processing systems. *Computational Linguistics*, *41*(4), 579-623. https://doi.org/10.1162/COLI_a_00232

[7] Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification*. Amsterdam: John Benjamins Publishing Company.

[8] Stefanowitsch, A. (2006). Words and their metaphors: A corpus-based approach. *Trends in Linguistics Studies and Monographs*, *171*, 63.