

A Technical Guide to the Metaphor Identification Procedure (MIPVU): From Theoretical Foundations to Empirical Corpus Analysis

Published: February 23, 2026 | Index ID: #115

Introduction to Systematic Metaphor Identification

Metaphor is no longer viewed merely as a decorative flourish of literary language or a rhetorical device reserved for poets and orators. Since the cognitive revolution in linguistics, sparked significantly by Lakoff and Johnson's seminal 1980 work, **Metaphors We Live By**, metaphor has been recognized as a fundamental cognitive mechanism. It is a tool through which humans understand abstract concepts by mapping them onto concrete, physical experiences. However, for decades, the study of metaphor faced a significant methodological hurdle: the lack of a reliable, transparent, and replicable system for identifying metaphors in natural language. This challenge led to the development of the **Metaphor Identification Procedure (MIP)** and its subsequent evolution into **MIPVU** (Metaphor Identification Procedure VU University Amsterdam).

The book "*A Method for Linguistic Metaphor Identification: From MIP to MIPVU*" by Gerard J. Steen and his colleagues represents a watershed moment in corpus linguistics. It provides a rigorous framework for moving beyond researcher intuition toward empirical data. By establishing an explicit manual of instructions, researchers can now identify metaphor at the level of **word use** with high inter-rater reliability. This article provides a comprehensive technical breakdown of these methodologies, their theoretical underpinnings, and their practical application in contemporary linguistic research.

The Theoretical Framework: Three-Dimensional Metaphor Analysis

Before diving into the procedural mechanics of MIPVU, it is essential to understand the three-dimensional perspective adopted by Steen. This framework distinguishes between three distinct but interrelated manifestations of metaphor:

- **Metaphor in Language:** This refers to the linguistic expression itself—the actual words used in a text or speech that can be traced back to a cross-domain mapping.
- **Metaphor in Thought:** This involves the underlying conceptual mappings (e.g., UNDERSTANDING IS GRASPING) that reside in the cognitive architecture of the speaker or listener.
- **Metaphor in Communication:** This dimension focuses on the deliberate use of metaphor as a tool for communication, where the metaphor is intended to be recognized as such by the addressee to achieve a specific communicative goal.

MIPVU primarily targets **Metaphor in Language**, providing the empirical baseline required to make further inferences about thought and communication. Without a reliable way to tag linguistic metaphors, any claims about conceptual metaphors remain speculative.

The Evolution: From MIP to MIPVU

The original Metaphor Identification Procedure (MIP) was developed by the **Pragglejaz Group** (2007) to provide a simple, reliable protocol. While effective, it had limitations, particularly in how it handled direct metaphors (like

similes) and implicit metaphors. MIPVU was developed to refine these gaps, offering a more granular approach for corpus-linguistic research.

Core Mechanics of the MIPVU Protocol

The protocol operates on the principle of **lexical units**. Rather than looking for "metaphorical ideas," the researcher examines every single word in a text. The process follows a strict algorithmic path:

1. Read the entire text: Establish a general understanding of the context.
2. Determine lexical units: Divide the text into individual words or multi-word expressions that function as single units.
3. Establish the contextual meaning: For each lexical unit, determine its specific meaning in the given context.
4. Check for a basic meaning: Consult a dictionary to see if the lexical unit has a more basic, contemporary meaning. Basic meanings are typically more concrete, related to bodily action, precise, or historically older.
5. Compare meanings: If the contextual meaning contrasts with the basic meaning but can be understood in terms of it, the word is marked as a Metaphor Related Word (MRW).

Technical Analysis: Identifying Metaphor-Related Words (MRW)

MIPVU introduces several technical categories to classify how a word functions metaphorically. This goes beyond the binary "metaphor/not metaphor" classification of the original MIP.

1. Indirect Metaphor (MRW-ind)

This is the most common type of metaphor. It occurs when there is a contrast between the contextual meaning and the basic meaning, and the contextual meaning can be understood through a comparison with the basic meaning. For example, in the sentence "The economy **fell** into a recession," the word "fell" is an indirect metaphor because its basic meaning involves physical movement downward due to gravity, which contrasts with its contextual meaning of a numerical or qualitative decrease.

2. Direct Metaphor (MRW-dir)

Direct metaphors occur when the metaphor is explicitly signaled, often through similes or analogies. In the phrase "The city is **like** a beehive," the comparison is explicit. MIPVU provides specific rules for tagging the signal (e.g., "like," "as") and the metaphorical vehicle.

3. Implicit Metaphor (MRW-impl)

These occur when a metaphor is implied by the presence of a pro-form or a substitution. For instance, if a speaker discusses a "path" (metaphor for life) and then says "I am still on **it**," the word "it" refers back to the metaphorical concept and is tagged as an implicit metaphor.

Comparison Matrix: MIP vs. MIPVU

The following table summarizes the key technical differences between the original Pragglejaz MIP and the expanded MIPVU framework.

Feature	MIP (Pragglejaz, 2007)	MIPVU (Steen et al., 2010)
Primary Focus	Word-level indirect metaphor	Comprehensive identification (Indirect, Direct, Implicit)
Unit of Analysis	Lexical Unit	Lexical Unit (with refined boundaries)
Direct Metaphor	Often excluded or ignored	Explicitly coded using signals (MRW-dir)
Polysemous Words	Broadly handled	Strict dictionary-based criteria (Macmillan/LDOCE)
Reliability Testing	Recommended	Mandatory (Cohen's Kappa / Fleiss' Kappa)
Application	General qualitative analysis	Large-scale quantitative corpus linguistics

The Dictionary-Based Approach: Defining "Basic Meaning"

A critical technical requirement of MIPVU is the use of a standardized dictionary. To prevent researcher bias, the **Macmillan English Dictionary** or the **Longman Dictionary of Contemporary English** are typically recommended. A meaning is considered "basic" if it meets the following criteria:

- **Concreteness:** Does the sense refer to something that can be perceived by the senses?
- **Bodily Action:** Is the sense related to physical movement or posture?
- **Precision:** Is the sense more specific than the contextual sense?
- **Historical Priority:** Is this the earliest recorded sense of the word (though this is often secondary to synchronically "basic" senses)?

If a word like "attack" is used in the context of a "verbal attack," the dictionary will reveal a basic sense involving physical violence. Because the verbal criticism can be understood in terms of physical violence, it is tagged as metaphorical.

Step-by-Step Field Guide for Researchers

Implementing MIPVU in a research project requires a disciplined workflow to ensure data integrity. Follow these steps for a successful corpus-linguistic analysis:

Phase 1: Corpus Preparation

1. **Select a representative corpus:** Ensure the text size is manageable for manual coding or that you have a trained team.
2. **Tokenization:** Use a software tool (like CLAWS or NLTK) to tokenize the text, but manually verify lexical units (e.g., phrasal verbs like "put up with" should be treated as one unit).

Phase 2: Training and Calibration

1. **Pilot Study:** Have multiple coders analyze a small subset of the text (approx. 500-1000 words).
2. **Reconciliation:** Discuss every disagreement. The goal is not just agreement, but alignment on why a word is or isn't a metaphor.

3. Calculation of Inter-rater Reliability: Use Cohen's Kappa. A score above 0.7 is generally required for scientific publication.

Phase 3: Formal Coding

1. Sequential Analysis: Code the text word by word. Avoid skipping ahead, as context is cumulative.
2. Documentation: Maintain a log of difficult cases. For example, the word "see" (understanding) is frequently debated; consistent rules must be applied across the entire corpus.

Troubleshooting and Common Challenges

Even with a rigorous manual, researchers encounter "grey areas." Here are common issues and their MIPVU-compliant solutions:

1. Dead Metaphors

Many words have metaphorical origins that are no longer felt by native speakers (e.g., "examine," which originally meant to weigh). MIPVU handles this by focusing on **synchronous** dictionary definitions. If the dictionary lists the concrete sense as a current, distinct sense, it is coded as metaphorical. If the concrete sense is obsolete, it is not.

2. Polysemy vs. Metaphor

Not all multiple meanings are metaphorical. Some are instances of metonymy or general polysemy. The key is the **Comparison** step. If the contextual meaning cannot be understood through a comparison with the basic meaning, it is not a metaphor.

3. Proper Names and Technical Jargon

Proper names (e.g., "The White House issued a statement") are usually excluded from metaphor analysis, as they function via metonymy. Technical jargon should be checked against general-purpose dictionaries; if the term is purely technical with no basic counterpart, it is not coded as MRW.

Case Study: English vs. Dutch Metaphor Use

The research by Steen et al. involved a cross-linguistic comparison between English and Dutch. By applying the same MIPVU criteria to both languages, the researchers were able to quantify the "metaphorical density" of different genres. They found that:

- Academic Texts: Show the highest density of metaphors (roughly 18% of all words), as abstract reasoning necessitates metaphorical scaffolding.
- Fiction: Surprisingly, fiction often has lower metaphor density than academic prose, though the metaphors used are often more "creative" or "novel."
- Conversation: Contains a high frequency of "fossilized" metaphors and idioms.

This cross-linguistic application proves that the MIPVU protocol is robust enough to handle the structural differences between Germanic languages, provided that reliable dictionaries are available for both.

Broader Implications for Linguistics and Beyond

The standardization of metaphor identification has profound implications for various fields. In **Applied Linguistics**, it allows for better syllabus design by identifying the metaphors that students must master to understand academic discourse. In **Political Science**, it enables the quantitative analysis of framing—how politicians use metaphorical language to shape public perception (e.g., "The War on Drugs").

Furthermore, MIPVU provides the necessary training data for **Natural Language Processing (NLP)**. Machine learning models for metaphor detection require high-quality, human-annotated datasets. By using MIPVU-tagged corpora, developers can train algorithms that recognize figurative language more accurately, improving sentiment analysis and machine translation.

In summary, the transition from intuitive metaphor spotting to the systematic application of MIPVU represents a maturation of the field. It allows metaphor research to move from the realm of literary criticism into the realm of hard social science. While the process is labor-intensive, the resulting data is defensible, replicable, and essential for any deep understanding of how language, thought, and communication intersect. Researchers adopting this method contribute to a global, converging body of evidence that continues to unravel the complexities of the human mind.

Tags: #MIPVU #Metaphor Identification #Corpus Linguistics #Cognitive Linguistics #Gerard Steen

