Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

CULTURE LETTERARIE E FILOLOGICHE

Ciclo 35

Settore Concorsuale: 01/B1 - INFORMATICA

Settore Scientifico Disciplinare: INF/01 - INFORMATICA

ETHICS IN THE FLESH: FORMALIZING MORAL VALUES IN EMBODIED COGNITION

Presentata da: Stefano De Giorgis

Coordinatore Dottorato

Marco Antonio Bazzocchi

Supervisore

Aldo Gangemi

Co-supervisori

Gilberto Corbellini
Elisabetta Sirgiovanni

Esame finale anno 2023

Abstract

Values are beliefs or principles that are deemed significant or desirable within a specific society or culture, serving as the fundamental underpinnings for ethical and socio-behavioral norms. The objective of this research is to explore the domain encompassing moral, cultural, and individual values. To achieve this, we employ an ontological approach to formally represent the semantic relations within the value domain. The theoretical framework employed adopts Fillmore's frame semantics, treating values as semantic frames. A value situation is thus characterized by the co-occurrence of specific semantic roles fulfilled within a given event or circumstance. Given the intricate semantics of values as abstract entities with high social capital, our investigation extends to two interconnected domains. The first domain is embodied cognition, specifically image schemas, which are cognitive patterns derived from sensorimotor experiences that shape our conceptualization of entities in the world. The second domain pertains to emotions, which are inherently intertwined with the realm of values. Consequently, our approach endeavors to formalize the semantics of values within an embodied cognition framework, recognizing values as emotional-laden semantic frames. The primary ontologies proposed in this work are: (i) ValueNet, an ontology network dedicated to the domain of values; (ii) ISAAC, the Image Schema Abstraction And Cognition ontology; and (iii) EmoNet, an ontology for theories of emotions. The knowledge formalization adheres to established modeling practices, including the reuse of semantic web resources such as WordNet, VerbNet, FrameNet, DBpedia, and alignment to foundational ontologies like DOLCE, as well as the utilization of Ontology Design Patterns. These ontological resources are operationalized through the development of a fully explainable frame-based detector capable of identifying values, emotions, and image schemas generating knowledge graphs from from natural language, leveraging the semantic dependencies of a sentence, and allowing non trivial higher layer knowledge inferences.

Keywords: Values, Image Schemas, Knowledge Representation, Frame Semantics

Acknowledgements

Thanks to my supervisor, Professor Aldo Gangemi, for being a constant source of inspiration as well as intensional and extensional knowledge.
Thanks to Professor Gilberto Corbellini and Professor Elisabetta Sirgiovanni for serving as co-supervisors.
Thanks to Professor Dagmar Gromann, for her expertise and helpfulness, without whose help part of this research could not have been accomplished.
Thanks to Professor Maria Hedblom, for her support, expertise, availability, for being an inspiration and for her inexhaustible energy.
Thanks to Fondazione Ruberti for funding this doctoral research.
Thanks to H2020 SPICE project for funding part of this research.

Contents

List of Figures

List of Tables

Introduction

In this work, we delve into the domain of moral and cultural values within the techno-social environment. To initiate our exploration of this domain, we must first address the question: “What is a value?” A moral or cultural value can be defined as a belief or principle that a particular society or culture considers important or desirable [274]. Values serve as the foundational knowledge for a society’s ethical and behavioral norms, and they are influenced by various factors such as religion, tradition, and personal experience, while also exerting influence over these aspects. Values are often reflected in the laws and customs of a society and play a role in shaping individuals’ interactions and decision-making processes. They reside in the intersection between the realms of moral philosophy and social psychology, possessing a dual nature: moral and socio-cultural.

Bilsky and Schwartz [25] describe values as structures that run parallel to social norms, distinguished by two main characteristics. Firstly, values lack explicit rules or formalization. Secondly, as Rozin [277] has argued, their system of reward and sanction operates on the emotional level. Bilsky and Schwartz individuate five features for values:

1. they are considered and treated as concepts or beliefs;
2. they inhere some desirable state of the world or behavior;
3. they can occur in specific situations, but they transcend them;
4. they are pivotal for selection or evaluation processes;
5. they are often organized by relative importance.

With the aim of formalizing the elusive concept of values, we leverage these fundamental features as the basis of our approach. Feature number 5 emphasizes the necessity of adopting an ontological structure to formally represent domain knowledge about values and enable automatic inferences. Feature number 3 indicates the need to represent entities that exist as encoded relational structures and manifest in our daily lives through the fulfillment of specific semantic roles. In line with Fillmore’s frame semantics [74], our approach treats the notion of value as a semantic frame, as defined below.

In this study, we incorporate key psychological and social theories that offer a taxonomic system for organizing values, including the Moral Foundations Theory [126], the Basic Human Values theory [288], and the Moral Molecules theory [42]. Through the adoption of the XD methodology [26] and the reuse of various Ontology Design Patterns [96], we transform these theories into ontological modules. By treating values as frames, we can model multiple theories concurrently and treat each theory as an independent ontological module. This approach is realized in the ValueNet ontology network, which aims to model the comprehensive concept of value along with its aforementioned nuances.

Within the ValueNet ontology network, the diverse interpretations of “value” proposed by different theories are represented as subframes of a generic notion of value. The generic notion serves as a superclass encompassing all the specific interpretations. This allows for the capture of different aspects of value semantics within distinct ontological modules, enabling the adoption of multiple theories simultaneously for automatic inferences.

Treating values as semantic frames prompts us to examine the structure of the frame and its lexicalization. In this work, we operationalize value theories by creating a knowledge base of semantic triggers, utilizing well-established semantic web resources such as WordNet [225], VerbNet [175], FrameNet [8], DBpedia [7], BabelNet [230], and ConceptNet [308], among others. These resources facilitate the representation of value-related concepts, enhancing the semantic richness of the ontological framework.

However, considering features 1, 2, and 4, the notion of value emerges as a complex social construct with semantics that are constantly negotiated between commonsense knowledge, individual perceptions of the world, moral ethics, and dynamics of emotional appraisal. To address these requirements, an expanded scope of inquiry becomes necessary. In response to feature 1, this work adopts an approach that conceptualizes values as embodied cognition. In addition to the ValueNet ontology, another ontology network introduced in this study is ISAAC, the Image Schema Abstraction And Cognition ontology.

Image Schemas, as defined by Johnson [159], serve as the fundamental building blocks of cognition. They are sensorimotor cognitive patterns derived from our bodily interactions with the environment, shaping our understanding of the world. Exploring how values are conceptualized starting from our embodied experiences becomes imperative in order to illuminate the intricate network of relationships and the ontological nature of values.

Furthermore, features 2 and 4 underscore the emotional dimension inherent in values. To account for this additional semantic facet, this work proposes another ontology network called EmoNet, which focuses on emotions.

As a result, this study takes the form of a transdisciplinary investigation into the concept of value. From this initial analysis, value can tentatively be defined as an “abstract entity with high social capital,” considering its multifaceted nature and its pervasive influence in

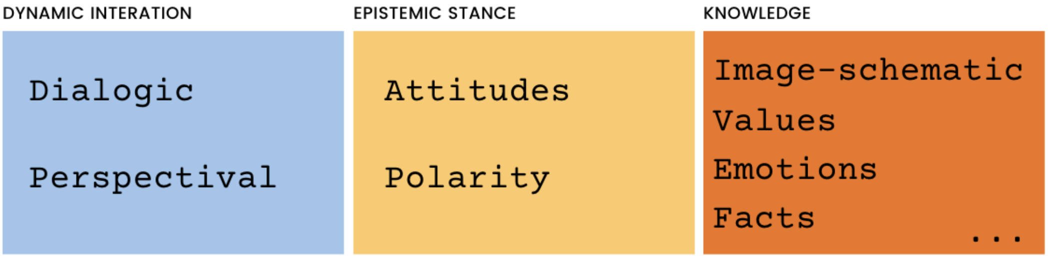


Figure 1: Commonsense Knowledge and Interaction.

various domains of human experience.

Aim of the Work

This study traverses multiple domains of knowledge and adopts a particular perspective to investigate the research conducted.

The orange box labeled "Knowledge" in Figure 1 depicts the layers of knowledge examined in this work. Starting from the bottom, there is factual knowledge of the world, encompassing data, information, and knowledge about various subjects. Above that, there are the emotion and value layers influenced by individual human interactions within the socio-cultural context. Finally, there is the image-schematic layer, which refers to the conceptualization shaped by an individual's mode of experiencing the world through their embodied physicality.

The yellow box represents the modes of expressing the content derived from the knowledge layers described in the orange box. Language, in particular, plays a significant role as a means of codified symbolic expression through verbalizing concepts, preferences, plans, and information. It constitutes an integral part of the expressive form of an individual's attitude towards a specific entity. We consider the combination of attitudes and attributed polarity as constituting the "epistemic stance" layer. Epistemic stance has been defined as the amalgamation of "attitudes, feelings, judgments, or commitments regarding the propositional content of a message" [24]. In this work, we specifically consider the expression of a cognizer's position and relationship with respect to an entity, which can be conveyed through self-connotations and/or entity connotations via the aforementioned higher layers of meaning. The perception of an entity as euphoric or dysphoric, representing pleasure or displeasure, determines the attribution of polarity. Understanding an individual's conceptual position entails representing the network of relations and meanings they create within a given context, both at conscious and unconscious levels. It involves comprehending their "perspective" on the world.

Finally, the blue box represents the dialogical interaction and exchange of knowledge packages derived from the aforementioned layers. This exchange is an integral part of any human interaction grounded in commonsense knowledge.

Ultimately, the primary objective of this research is to illuminate the intricate interplay between explicit and, more notably, implicit connections among values, emotions, and embodied cognition. The aim of this study is to enhance our understanding of how external phenomena in the world are perceived through sensori-motor interactions, internalized, categorized as instances of specific schemas, and subsequently evaluated based on particular value and emotion profiles.

Relevance of the Work

To demonstrate the relevance and timeliness of this project, it is pertinent to refer to the guidelines outlined in the Berlin Declaration on Digital Society and Value-Based Digital Government [1] in 2020, as well as the European Union Guidelines for Trustworthy AI[2] released in 2019. By considering these criteria, we offer the following comments to demonstrate the coherence and consistency of this work, particularly with the European Union AI Guidelines:

Here are some comments to demonstrate how this work aligns with the criteria outlined in the European Union AI Guidelines:

1. *Human agency and oversight*: By researching human and cultural values, this work delves into the investigation of both innate human cognition features and culturally-dependent social norms. Making latent moral content related to fundamental human values explicit can enhance informed decision-making, thereby empowering individuals.

2. *Technical Robustness and safety*: The approach adopted in this work, which is based on knowledge representation, ensures robustness through the utilization of automatic inference reasoners. Moreover, it provides control over the knowledge base and the axiomatic layer, promoting accuracy, reliability, and reproducibility.

3. *Privacy and data governance*: This research extensively utilizes open peer-to-peer published resources and, when incorporating data containing sensitive information, thoroughly anonymizes the data. This approach prioritizes full respect for privacy and data protection, while adhering to adequate data governance mechanisms.

[1]The Declaration is available here:https://digital-strategy.ec.europa.eu/en/news/berlin-declaration-digital-society-and-value-based-digital-government

[2]A partial excerpt is provided here for brevity, the complete digital strategy bullet list is accessible at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

4. *Transparency*: The reuse of well-known semantic web technologies aligns with the principle of transparency. Additionally, all the developed ontologies and the produced data from experiments are available and queryable online through a stable endpoint, promoting transparency and traceability.

5. *Diversity, non-discrimination, and fairness*: The investigation of human values in this work addresses the issue of unfair bias, aiming to avoid negative implications such as the marginalization of vulnerable groups, prejudice, and discrimination. By detecting latent moral content, the system can raise awareness of potential ethical concerns and contribute to mitigating cognitive and cultural biases.

6. *Societal and environmental well-being*: The ability to automatically detect the presence or absence of a universalistic vision as a value aligns with the objective of benefiting all human beings, including future generations. This work contributes to the broader goal of ensuring AI systems are sustainable and environmentally friendly.

7. *Accountability*: The developed detection and reasoning pipeline in this work is explainable, well-documented, and grounded in established theoretical literature. These aspects contribute to the establishment of mechanisms for responsibility and accountability for AI systems and their outcomes.

By addressing these key aspects of the European Union AI Guidelines, this research demonstrates its consistency and adherence to the outlined criteria.

Research Questions

This work is interdisciplinary not *de dicto*, namely due to the interdisciplinarity of the themes faced, but *de re*, meaning it encompasses pragmatically multiple disciplines rather than just referring to them conceptually. The investigation of "values" as a topic has been explored in various fields, including philosophical ethics, economics, anthropology, psychology, game theory, scopistic theory [37], and sociology, each shedding light on different aspects and interpretations of values.

The primary objective of this research is to examine the ontological nature of moral and cultural values using state-of-the-art frameworks and AI techniques for knowledge representation and extraction. Additionally, the operationalization of theoretical frameworks to detect values from natural language necessitates the exploration of two intertwined domains: embodied cognition and emotions.

This work is guided by the following research questions:

- **RQ1**: What is a value, according to a certain theory?

  - *RQ1.1*: What are the main theories formalising the domain of moral values?
  - *RQ1.2*: What are the possible inferences of a certain theory?

- **RQ2**: How are values conceived, and consequently conveyed in natural language?
  - *RQ2.1*: Is it possible to detect values from natural language with knowledge-based methodology?
  - *RQ2.2*: If possible, how are values conveyed lexically and what is the degree of their embodiment?
  - *RQ2.3*: What is the relation between values and emotions?

By addressing these research questions, this work aims to deepen our understanding of values, their theoretical foundations, their manifestation in natural language, and their connection to emotions and embodied cognition.

This work acknowledges that the domains of values, emotions, and image schemas are vast and complex, encompassing various perspectives and avenues of investigation. While this research does not claim to exhaustively address all possible inquiries within these domains, its primary objective is to demonstrate the interconnectedness and interplay of these layers of meaning. By delving into the ontological nature of values, exploring the role of emotions, and investigating the embodiment of image schemas, this work seeks to establish a solid foundation for future research endeavors. It aims to provide valuable insights into the intricate relationships between these domains, paving the way for further exploration and deeper understanding in the field.

Thesis Outline

Chapter 1 This chapter begins by providing definitions of key notions for this work, namely embodied cognition, frame semantics, and ontological structure. It also outlines the methodology used for ontology modeling, with a specific focus on the Framester ontology hub. The chapter further describes the process of operationalizing modeled theories through the identification of semantic triggers and the reuse of entities from other semantic web resources. Lastly, it delves into the practices employed for ontology testing, which will be applied in subsequent chapters.

Chapter 2 This chapter centers on the domain of image schemas. It starts by presenting a state-of-the-art overview and then introduces the ISAAC ontology network. Within ISAAC, four modules are dedicated to image-schematic theories, while two modules address cognitive metaphors. The ontology testing section of this chapter encompasses the description of an automatic, frame-based image schema detector designed to identify image schemas from natural language.

Chapter 3 The core focus of this chapter is values. It begins with an explanation of the concept of value in various fields and provides an overview of the main theories from social and moral psychology. Subsequently, the ValueNet ontology is introduced, incorporating key theories such as the Moral Foundations Theory, Basic Human Values theory, and Moral Molecules theory. Additionally, an extra ontology module is created to include individual notions of value extracted from web sources, capturing everyday behaviors not covered by traditional value theories. The ontology testing section involves automatic inferences derived from populating the aforementioned ontology modules, as well as the development of a frame-based value detector capable of recognizing the evocation of value frames from the aforementioned theories.

Chapter 4 This chapter offers a concise overview of emotion theories and introduces the EmoNet ontology, with a particular focus on Ekman's Basic Emotions theory. The chapter concludes with a section on ontology testing and the development of a frame-based detector for automatic extraction of emotions from natural language.

These chapters collectively contribute to the exploration of embodied cognition, image schemas, values, and emotions, offering a comprehensive understanding of their ontological nature, semantic triggers, and detection mechanisms.

Chapter 5 Lastly, an experiment is conducted to extract values, emotions, and image schemas from natural language by combining the three detectors presented in the previous chapters. The results are analyzed, and patterns of semantic dependencies and co-occurrences among the three layers are identified. A comprehensive discussion concludes the chapter.

The final Conclusions section of the thesis provides a summary of the achievements and proposes potential future directions for further developments.

Chapter 1

Fast and Framal Toolkit

This chapter is divided into two parts. The first part provides a brief introduction to some theoretical preliminaries, while the second part describes the tools and resources necessary to comprehend the ontology building process discussed in the subsequent chapters.

Section 1.1 presents an introduction to the concepts of "Grounding," "Frame," and "Ontology" with minimal theoretical background. These notions are extensively utilized throughout this work. It furthermore introduces the fundamentals of frame semantics.
Section 1.2 focuses on the methodology employed to model the ontologies presented in the following chapters. It presents the Framester ontology as well as several semantic web resources reused in this work.
Section 1.3 describe some modeling pitfalls and necessary trade-offs when modeling higher layers of meaning.
Section 1.4 poses some epistemological modeling dilemmas and describe the Exuviae ontological module.
Section 1.5 outlines the process of knowledge extraction from semantic web resources, which is utilized to populate the semantic trigger graphs for each value, emotion, and image schema, as elaborated in subsequent chapters.
Lastly, Section 1.6 centers on the existing FRED tool and the FRED-based methodology employed to generate knowledge graphs from natural language. These graphs are subsequently enriched with image schema, value, and emotion knowledge.

1.1 Grounding, Frames, and Ontologies

The title of this chapter draws inspiration from Gigerenzer's "Fast and Frugal Toolkit" [106], which refers to a set of heuristics enabling humans to make adaptive choices quickly in decision-making scenarios within real-world environments. Similarly, in this section, we provide the theoretical and practical tools necessary to comprehend the subsequent chapters.

1.1.1 Grounding: Cognitive Aspects of Meaning

The term "grounding" is used here to refer to grounded cognition. Grounded cognition has been influenced by various disciplines, including philosophy, perception, cognitive linguistics, psycholinguistics, cognitive psychology, and cognitive neuropsychology. The central assumption of grounded cognition is that cognition is rooted in bodily states and situated action. Cognitive linguists have proposed two key concepts: mental spaces, which are possible worlds based on experience expressed in truth-conditional semantics, proposed in particular by Fauconnier [71], and cognitive grammars, which encompass the continuum of grammar, semantics, and lexicon, proposed in particular by Talmy [319] and Langacker [192] as explanations for language and the mind. The conceptual metaphor hypothesis put forth by Lakoff and Johnson in [190], posits that physiological experience serves as the foundation for abstract conceptualizations. In cognitive psychology, Barsalou [15, 14], asserts that knowledge is built upon a compositional system of perceptual symbols. Moreover, Violi [337] discusses different senses of "embodiment.":

- The first sense, included in the Enaction paradigm, suggests that every cognitive process is carried out by a material entity;
- The second sense posits that for a cognitive process to exist, it must be enacted by a corresponding entity;
- The third sense, proposed by Lakoff and Johnson [159], views embodiment as a necessary starting point, considering material constraints, from which theoretical hypotheses can be derived.

The process of categorizing the realm of reality serves as both an inherent aspect of human cognition and a consequence of our neurophysiological makeup. This work embraces this idea by recognizing that every conceptualization is a result of prior perception of the world. This perspective is crucial for the development of AI systems as it acknowledges the need to consider not only the internal processes of these systems but also their interaction with the environment. This understanding is particularly important for the advancement of cognitively inspired AI systems, such as robots or virtual agents, that operate in complex and dynamic settings. Although the research conducted in this context does not primarily focus on the neuroscientific aspects of embodied cognition and moral values, it acknowledges that values are social constructs resulting from negotiated conceptualizations between individuals and society, and their conceptualization is significantly influenced by embodiment.

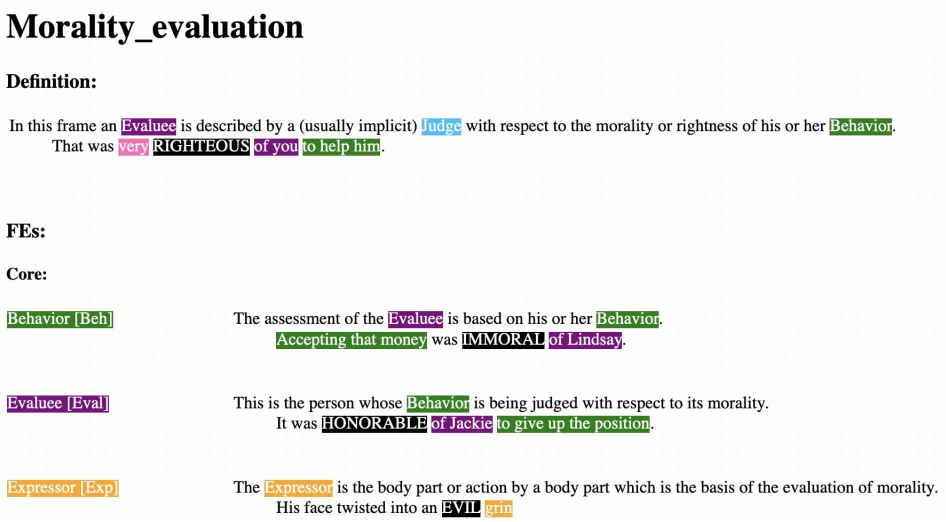
An illustrative example of this relationship can be found in Conceptual Metaphor theory, and the MetaNet repository [63] (described in Section 1.2.2). The Conceptual Metaphor theory describe that process of understanding or describing a certain physical or abstract domain in terms of another. MetaNet is a repository that contains numerous conceptual metaphors that shape our everyday cognition. More than 20 conceptual metaphors directly pertain to the concept of "morality," such as `MORAL CORRUPTION IS A DESTRUCTIVE FORCE` or `MORALITY IS PURITY`. Additionally, several metaphors are related to frameworks based on specific values, such as `DEMOCRACY IS AN EQUAL PARTNERSHIP` or `TRUST RELATIONSHIPS ARE BUILDINGS`. Many of these metaphors associated with morality are image-schematic, such as `MORALITY IS UP`, `IMMORALITY IS DOWN`, and `MORALITY IS A STRAIGHT PATH`.

In essence, this work adopts a grounded cognition approach to explore the interconnectedness between perception, action, and the cognitive aspects of values.

1.1.2 Frames and Frame Semantics

The main theoretical foundation of this work is derived directly from Fillmore's frame semantics [74]. Frame semantics has proven effective in bridging the gap between the encyclopedic nature of language and its linguistic expression, while embodied cognition emphasizes the importance of contextualizing this relationship within an embodied dimension. A critical aspect of frame semantics, which is also adopted in this work, is the notion that the meaning of something, such as a segment of reality associated with a lexical unit, encompasses all the encyclopedic knowledge associated with that entity. Consequently, the semantics of a specific lexical unit cannot be fully comprehended without accessing the essential knowledge that connects other parts or aspects of the world to that lexical unit. As Fillmore [75] aptly explains:

> Semantic theories founded on the notion of cognitive frames or knowledge schemata, by contrast, approach the description of lexical meaning in a quite different way. In such theories, the word's meaning can be understood only with reference to a structured background of experience, beliefs, or practices, constituting a kind of conceptual prerequisite for understanding the meaning. Speakers can be said to know the meaning of the word only by first understanding the background frames that motivated the concept that the word encodes. Within such an approach, words and word senses are not related to each other directly, word to word, but only by way of the links to common background frames and indications of the manner in which their meanings highlight particular elements of such frames.



Morality_evaluation

Definition:

In this frame an Evaluee is described by a (usually implicit) Judge with respect to the morality or rightness of his or her Behavior.
That was very RIGHTEOUS of you to help him.

FEs:

Core:

Behavior [Beh]	The assessment of the Evaluee is based on his or her Behavior. Accepting that money was IMMORAL of Lindsay.
Evaluee [Eval]	This is the person whose Behavior is being judged with respect to its morality. It was HONORABLE of Jackie to give up the position.
Expressor [Exp]	The Expressor is the body part or action by a body part which is the basis of the evaluation of morality. His face twisted into an EVIL grin

Figure 1.1: The Morality_evaluation frame, according to FrameNet resource.

In a broad definition, a frame is a cognitive representations of prototypical features of a situation. Frames are structures that formalize the network of meaning through semantic roles that participate in a given situation. This network of semantic references to triggers of meaning is commonly referred to as the “activation” or “evocation” of a frame.

Figure 1.1 illustrates the schematic representation of semantic relations for the `fn:Morality_evaluation` frame in FrameNet (the FrameNet resource is described in Section 1.2.2). In the FrameNet resource, semantic roles are referred to as “frame elements” (FEs). The “Core” frame elements are essential for the realization of the frame, representing roles that must be filled to realize a certain type of situation.

It is important to note that the example presented in Figure 1.1 is provided here solely to demonstrate the composition of a frame and its roles. However, our conviction, which also motivates this work, is that this representation of a “morality evaluation” is highly incomplete and simplistic.

Frames, as a schematic representation of experience, bear resemblance to other schemas introduced in the field of AI, such as the concept of *scripts* introduced by Schank and Abelson [283]. Fillmore explicitly compares frames to other notions, such as experiential gestalt [190], asserting that frames can encompass a unified framework of knowledge or a coherent schematization of experience. Therefore, widely recognized frames offer a theoretically well-founded and practically validated basis for common-sense knowledge patterns.

The term “frame” itself has been employed by Minsky [226], Winograd [347], Bartlett [16], Langacker [192], and, of course, Lakoff and Johnson [187]. In the field of AI, frame semantics has been instrumental in developing more sophisticated and nuanced models of natural language understanding. Fillmore’s frame semantics, in particular, has had a significant impact by structuring the combination of linguistic descriptors and features of related knowledge structures to describe cognitive phenomena. Lexical units and sentences are semantically associated via frames, for their being schematic structures organising the common scenes evoked by a certain lexical unit. Traditional approaches to natural language processing often rely on fixed sets of rules or representations to interpret the meaning of words and sentences. In contrast, frame semantics-based approaches allow for more flexible and context-dependent interpretations of language. These approaches have been employed in AI to enhance the ability of AI systems to comprehend the relationships between different lexical units and concepts.

The FrameNet resource can be regarded as the optimal means of integrating cognitive and computational semantics, providing a structured framework for connecting linguistic material with encyclopedic and factual knowledge.

In this work, frame semantics is adopted as a formal structure for organizing semantic relations within the knowledge domain of values, as well as between the domain of values and image schemas and emotions. The following section outlines the utilization of ontologies, semantic web, and the practices employed in this work to formally represent knowledge in ontological structures.

1.1.3 Ontologies and Semantic Web Resources

In philosophy, the original meaning of "ontology" comes from the Greek term ὤν ὄντος, which refers to "the essence of being." The early inquiries of this discipline explored the fundamental nature of the world and investigated relationships within the realm of reality. Although the term "ontology" has a long and semantically significant history, in the context of this work, it is used solely within the domains of computer science and AI, as described below.

Uschold [331] defines ontology as "a model that represents a particular subject matter." Tom Gruber [117], in his work "What is an Ontology," offers a more comprehensive definition:

> In the context of computer and information sciences, an ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application.

Guarino [118] extensively elaborated on Gruber's definition, in particular focusing on ontological commitment and conceptualizations. For further elaboration on the notion of ontology as a document enabling a formal representation of a state of affairs in a possible world, and for a brief history of the term “ontology” in Information Science, refer to Neuhaus work [231].

While we acknowledge the ongoing debate regarding the notion of ontology itself, and the distinction between ontology and knowledge graph, providing a clear differentiation between the two is beyond the scope of this work. Henceforth, when used, the terms will carry the following meanings:

- Ontology: Refers to the semantic data model that represents a specific domain or ontological module.
- Knowledge Graph: Refers to data organized in a graph structure, where the schema relies on one or more ontologies.

To elucidate the usage distinction in this work, the transposition of any theory concerning image schemas, values, or emotions into formal semantics is referred to as an ontological module. On the other hand, the dataset organized in a graph structure, containing e.g. lexical and factual triggers for specific image schemas, values, or emotions, without providing a data model but rather a mere list of assertions, is termed a knowledge graph.

The methodology employed in this work gives rise to several questions that need to be addressed independently of the specific domain under examination. These questions stem directly from the use of the ontological structure and include:

- **Q1**: What entities exist within the domain?
- **Q2**: What are the relationships among the entities in the domain?
- **Q3**: What knowledge can be inferred through automatic reasoning on the model representing the domain?

These questions are inherent to the structure adopted in this work, and thus, they are reiterated in each chapter dedicated to a specific domain during the description of the ontological module and in the section on ontology testing.

Furthermore, the objective of this work extends beyond formalizing the value domain from the aforementioned perspectives; it aims to integrate within the broader context of semantic web resources. The term “Semantic Web,” coined by Tim Berners-Lee, the inventor of the World Wide Web, refers to a web that can be processed by machines, wherein much of the meaning is machine-readable [19]. The concept of the semantic web is based on the sharing of more than just raw strings; it involves the sharing of genuine encyclopedic knowledge.

To align this work with the interconnection of semantic web resources, we adopt a graph structure, specifically RDF graphs. RDF graphs consist of labeled entities and data values. Entities are expressed as Internationalized Resource Identifiers (IRIs), while data values can be literals. Blank nodes are denoted by scope identifiers (e.g., _:node1). The structure of an RDF graph can be represented as a set of triples, denoted as *(s, p, o)*, where each triple comprises a subject s, a predicate p, and an object o. A triple signifies the existence of a relation p between the subject s and the object o within a given universe of discourse Δ^i.

The Web Ontology Language (OWL), specifically OWL 2, is the chosen language for modeling knowledge domains. For a comprehensive description of OWL 2 and RDF, we refer to the documentation provided by the World Wide Web Consortium[1] (W3C).

Having introduced the theoretical framework that underlies this work, we will now proceed to describe the methodology employed for modeling the domains of image schemas, values, and emotions. Additionally, we will discuss the key semantic web resources that were utilized, as well as the technical tools employed for both modeling and testing the ontological modules.

[1] The OWL 2 documentation is available here: `https://www.w3.org/TR/owl2-overview/`

1.2 Ontology Building

This section focuses on the frameworks, working pipeline, and methodologies employed in the development of the ontological modules presented in the subsequent chapters. In Section 1.2.1, we introduce the eXtreme Design methodology utilized for ontology modeling, along with the NeOn guidelines that were adopted to ensure adherence to good modeling practices and ontology design patterns. Section 1.2.2 provides an overview of the Framester ontological hub and highlights the main semantic web resources that were reused in this work.

Following the description of these resources, Section 1.3 addresses the potential fallacies and incompletenesses that this work may encounter. This discussion serves two purposes: firstly, to elucidate the complexity, both technical and conceptual, involved in formalizing the aforementioned domains, and secondly, to summarize the iterative modeling process employed for all the ontological modules developed in this work.

In Section 1.4, we introduce the "meta-module" used to represent formal theoretical dependencies between models. Then, in Section 1.5, we present the QUOKKA workflow, providing a detailed account of how the theoretical ontological modules are operationalized in dedicated knowledge graphs. These knowledge graphs leverage semantic web resources aligned with the Framester ontology to retrieve image-schematic/value/emotion triggers.

1.2.1 eXtreme Design (XD) Methodology

The XD methodology [26] was initially introduced as part of the NeOn methodology [315, 316], but it can also be considered a standalone method for ontology engineering. Drawing inspiration from the eXtreme Programming (XP) agile software methodology, XD encompasses steps for project initiation, requirements analysis, development, testing, and release of an ontology.

An essential aspect of XD is the involvement of domain experts and their feedback. In the context of this work, which involves the transposition of sociological, psychological, and cognitive theories into ontologies, domain experts play a crucial role in faithfully translating the theories and engaging in direct discussions with the original authors whenever possible.

The XD methodology is described as "task-focused," meaning that ontologies built using XD are specifically designed for a set of targeted tasks rather than being generic representations of a knowledge domain. The methodology is iterative, with each iteration building upon the previous one, resulting in a modular final product that is easily reusable.

The workflow of the XD methodology can be summarized as follows:

1. Identify the requirements for the ontology design process and formulate a set of competency questions (CQ) that the ontology should be able to answer;

2. Assess whether existing Ontology Design Patterns (ODPs) from the Content ODPs repository[2] meet the modeling requirements and can be reused;

3. Validate the ontology modules through error provocation, inference testing, and validation for each module in the ontology network;

4. Integrate the modeled and tested ontologies into a closure module and populate them with domain entities from knowledge graphs.

The XD methodology is applied to each of the ontological modules developed in this work.

Ontology Design Patterns Ontology Design Patterns (ODPs) [254, 130] provide a solution to the challenges associated with reusing large ontologies. ODPs are designed to be smaller, more adaptable "building blocks" compared to entire ontologies. Their purpose is to facilitate the reuse of ontological components and promote interoperability among different ontologies.

The utilization of ODPs can vary. They can be directly incorporated into ontologies by using the OWL building components provided by the ODPs. Alternatively, they can serve as a source of inspiration and a conceptual framework for ontology development.

In this work, two ODPs, namely *Description&Situation* and *Agent-Role*, are extensively reused in the proposed ontological modules. These ODPs have been selected for their relevance to the specific domains being modeled. The complete repository of ODPs is available on the ODPs website[3], which provides a comprehensive collection of reusable patterns for ontology engineering.

In the next section, when describing the Framester ontology, these ODPs will be further discussed and their application in the ontological modules will be explained in detail.

1.2.2 Framester Ontology

In this section, we introduce the Framester ontological hub, which serves as framework for formal frame semantics in OWL 2. The main objective of the Framester hub is to establish interoperability among various lexical and factual resources that have been re-engineered as knowledge graphs or directly reused, while aligning them with frames and foundational ontologies. The architecture of Framester enables inheritance and unification within the

[2] The Content ODPs repository is available at http://ontologydesignpatterns.org/wiki/Submissions:ContentOPs

[3] The ODPs repository is available at http://ontologydesignpatterns.org/wiki/Main_Page

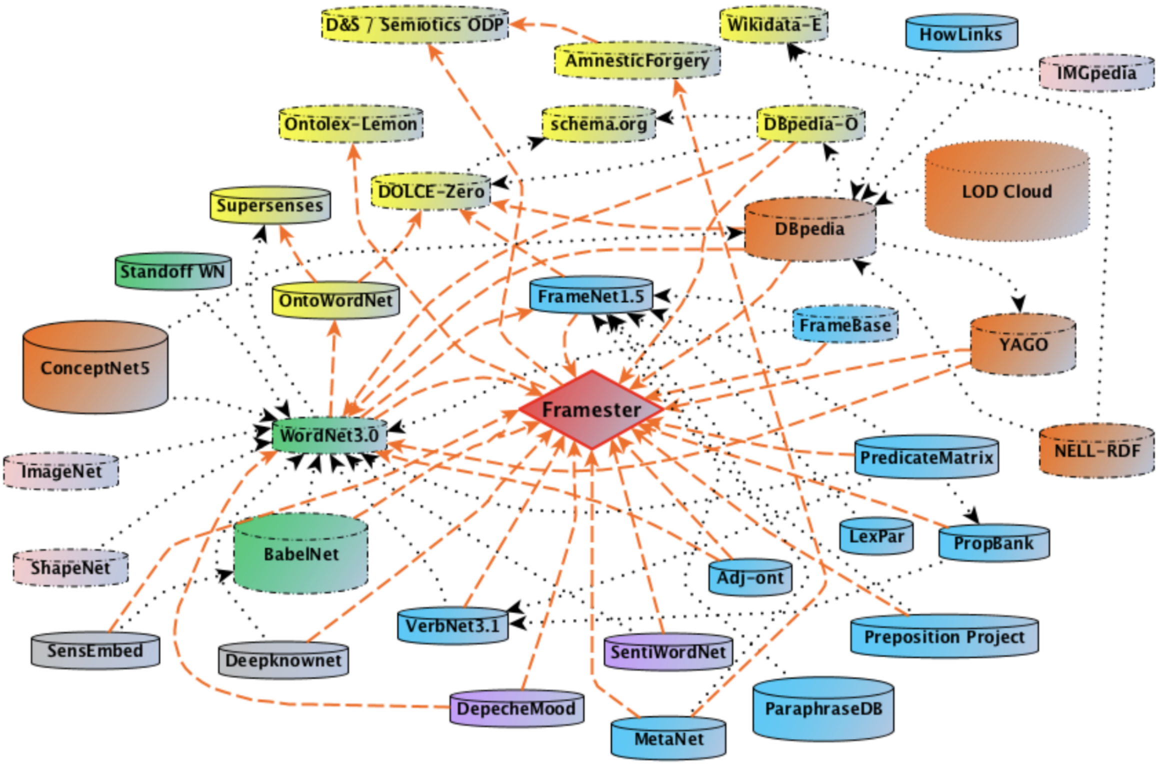


Figure 1.2: The Framester Ontology Hub

integrated resources, facilitating a comprehensive and interconnected knowledge representation.

Framester, as described in previous studies [90, 89], offers a formal semantics for frames by curating linked data versions of multiple linguistic resources. To establish connections between our ontologies and linguistic examples, we rely on formal representations of frames extracted from FrameNet [236]. In the following paragraphs, we provide a brief overview of the key resources integrated within Framester.

Among the integrated resources, we have FrameNet, WordNet [225], VerbNet [286], and a cognitive layer that includes MetaNet [91]. Additionally, multilingual resources like BabelNet [230], factual knowledge bases such as DBpedia [7] and YAGO [317], and foundational ontologies like DOLCE-Zero [94] are incorporated. These resources are interconnected through formal links, resulting in a cohesive RDF/OWL knowledge graph that encompasses diverse domains and facilitates cross-referencing and knowledge integration.

Framester can be used to jointly query (via a SPARQL endpoint[4]) all the resources aligned to its formal frame ontology[5].

[4] `http://etna.istc.cnr.it/framester2/sparql`
[5] The Framester Schema is available at: `https://w3id.org/framester/schema/`

In addition to the aforementioned resources, we incorporate three key knowledge layers in this work. Firstly, a sensori-motor cognitive layer comprising ISAAC [53] and ImageSchemaNet [52], which are presented in Chapter 2. Secondly, a moral and cultural values knowledge layer formalized in the ValueNet module [51, 6], described in Chapter 3, and thirdly, an emotion knowledge layer formalized as EmoNet, detailed in Chapter 4.

Within the framework of Framester semantics, image schemas, values, and emotions are treated as frames. They are modeled as frames, encompassing semantic roles to be filled, establishing relationships with other frames within FrameNet, and being evoked by lexical units. It is worth noting that in the context of FrameNet, frames are also referred to as *situation types*. Consequently, in the context of Framester semantics, observed, recalled, anticipated, and imagined situations are considered occurrences of frames.

FrameNet FrameNet, built upon the theory of meaning known as Frame Semantics developed by Fillmore and his colleagues, forms the basis of the FrameNet project. As elaborated in Section 1.1.2, the core idea behind Frame Semantics is that the meaning of words can be understood in terms of a semantic frame. A semantic frame provides a description of a specific type of event, relation, or entity, along with its associated actors.

To illustrate this concept, let's consider the example of cooking. The concept of cooking typically involves several essential roles: a cook, the food being prepared, possibly a container to hold the food while it is being cooked, and a heating source.

In the FrameNet ontology [236], which offers a formal representation of Fillmore's frame semantics, frames are also described as *situation types*. In the context of Framester semantics [89], observed, recalled, anticipated, and imagined situations are therefore regarded as occurrences of frames.

In this work, each image schema, value, or emotion is treated as a frame, while its manifestation in a specific context is considered a frame occurrence. The semantic roles played by the frame elements are represented by reusing the Agent-Role Ontology Design Pattern for each frame under consideration.

WordNet The WordNet project [225] is a linguistic resource that originated in the 1990s with the aim of providing a psycholinguistic resource for English based on the application of the linguistic Relational principle. Over the past 20 years, WordNet has expanded to cover multiple languages through the efforts of the GlobalWordNet association.

In WordNet, cognitive synonyms are organized into sets called "synsets," which consist of nouns, verbs, adverbs, and adjectives that express nearly identical meanings in a specific context (they are contextual synonyms). For instance, the synset `wn:happiness-noun-1` represents the semantic concept of "happiness. The synset encompasses words such as "happiness," "joy," "bliss," "jubilance," and others that share similar contextual meanings.

While WordNet may resemble a thesaurus in terms of word classification based on meaning, there are significant differences. Firstly, WordNet establishes connections between distinct word senses, going beyond mere word forms or strings of letters. As a result, words located in proximity to each other in the network are semantically related. Secondly, WordNet annotates semantic relationships between words, whereas a thesaurus primarily groups words together based on similarity in meaning. The semantic relationships captured by WordNet include synonymy, hyperonymy (superset), hyponymy (subset), meronymy (part-whole relation), and type (which distinguishes between types of entities, such as "furniture", and individuals, such as "Italy").

PropBank PropBank [261] is a corpus that annotates the semantic roles linking each argument of a predicate in a frame-like structure. These semantic roles are referred to as "arguments" in PropBank. The annotations in PropBank are built upon the phrase structure annotation of the Penn TreeBank [216]. The annotation process in PropBank involves assigning a sense ID (also known as a frameset or roleset ID) to each predicate, in addition to annotating its semantic roles. Consequently, for every verb in each tree representing a sentence's phrase structure, a PropBank instance is created, containing the sense ID of the predicate and its arguments annotated with semantic roles.

VerbNet VerbNet, developed by Schuler [286], is the largest online network of English verbs that establishes connections between syntactic and semantic patterns. It serves as a comprehensive verb lexicon and is hierarchically organized, with mappings to WordNet, PropBank, and FrameNet. VerbNet enhances Levin classes [196] by refining and extending them through the inclusion of subclasses, ensuring syntactic and semantic consistency among class members. Each verb class in VerbNet is fully characterized by thematic roles, selectional preferences of the arguments, and frames composed of a syntactic description and a semantic representation. The semantic representation includes subevent structure, following the Dynamic Event Model proposed by Pustejovsky [259] and Moszkowicz [262].

DBpedia DBpedia [7] is an initiative that leverages crowdsourcing to extract structured content from data generated by various Wikimedia projects. This curated data forms an open knowledge graph (OKG) accessible on the web. The DBpedia knowledge base offers several advantages compared to other knowledge bases: it is multilingual, covering a wide range of topics, and represents a consensus among the community regarding common knowledge. Utilizing the DBpedia knowledge base, one can pose intriguing queries based on factual knowledge, such as 'Provide a list of German philosophers born in the 18th century' or 'Give me information on Italian intellectuals who were exiled.'

The DBpedia knowledge base finds applications in diverse fields, ranging from enterprise knowledge management to reimagining search functionalities within Wikipedia.

Wikidata Wikidata is described as a 'free, collaborative, multilingual, secondary database' that collects structured data to support Wikipedia, Wikimedia Commons, and other wikis within the Wikimedia movement, as well as individuals worldwide. It is published under the Creative Commons Public Domain Dedication license, making it freely available. Wikidata operates as a collaborative platform maintained by Wikidata editors and acts as a secondary database that includes metadata about statements and facts.

BabelNet BabelNet [230] is a comprehensive multilingual encyclopedic dictionary that establishes connections among concepts and named entities through a vast network of semantic relationships. It offers extensive coverage of lexicographic and encyclopedic information for various terms. BabelNet expands upon the WordNet model, which is based on synsets, by incorporating multilingual lexicalizations. Each BabelNet synset represents a specific meaning and includes synonyms expressing that meaning in multiple languages.

YAGO YAGO [317] is a prominent knowledge base within the Linked Open Data cloud. It unifies Wikidata and schema.org into a consistent ontology that enables semantic reasoning using OWL 2 description logics.

ConceptNet ConceptNet [308] is a multilingual knowledge base that represents everyday words and phrases along with the meaningful relationships between them, designed to align with human understanding. The information in ConceptNet is collected from various sources, including expert-created resources, crowd-sourced contributions, and games with a purpose. The ConceptNet schema encompasses several relations such as `cn:DerivedFrom` and `cn:EtymologicallyDerivedFrom` (origin), `cn:Causes` (teleological), `cn:Antonym` (opposite meaning), `cn:IsA` (type), `cn:EtymologicallyRelatedTo`, `cn:SymbolOf`, `cn:UsedFor` (functional), `cn:HasSubevent`, `cn:MotivatedByGoal` (telic), and `cn:FormOf`.

These relations are utilized in the QUOKKA workflow for populating knowledge graphs, as described in Section 1.5.

DOLCE Foundational Ontology The Descriptive Ontology for Linguistics and Cognitive Engineering (DOLCE) [27, 94] is a top-level foundational ontology that finds wide application across various disciplines. An extensive investigation about the nature of foundational ontologies and their main differences can be found in Keet's work [167], as well as on her online material[6]. Foundational ontologies provide a high level categorization about classes of entities of the world, formalizing concepts such as 'Event,' 'Process,' relations such as 'Partonomy,' etc. Drawing inspiration from cognitive and linguistic theories, DOLCE aims to model commonsense knowledge and the understanding of reality as employed by individuals in diverse contexts, including socio-technical systems, manufacturing, financial transactions, and cultural heritage. DOLCE builds upon well-established ontological approaches, such as OntoClean [119], and relies on explicitly expressed philosophical concepts while clearly stating its ontological choices. With these characteristics, DOLCE has had a significant influence on many of the currently used top-level ontologies and has been employed to develop or enhance standards and publicly accessible resources, including CIDOC CRM [32], DBpedia, and WordNet. As a foundational ontology, DOLCE does not focus on any specific domain knowledge. Instead, it provides the general categories and relations necessary to present a coherent perspective of reality, integrate domain knowledge, and mediate between different domains. Over nearly two decades, DOLCE has demonstrated its ability to achieve stability in applied ontologies, and it has shown that interoperability between reference and domain ontologies is achievable. All the ontological models in this work are aligned with a specific version of DOLCE known as DUL, which stands for DOLCE Ultralight[7]. This version incorporates a constructionist view introduced by the Description/Situation Ontology Design Pattern, which will be described in the following paragraph.

Description&Situation ODP The DnS pattern [95] for "meta-level/syntactic sugar" serves as an extension to the DOLCE foundational ontology. Its purpose is to enable the predication of attributes concerning contexts, methods, norms, theories, situations, and models at the first-order, allowing for a partial specification of these entities. For instance, when considering values, any value frame can be represented as a reified N-ary relation, which is an intensional `dul:Description` satisfied by an extensional `dul:Situation` (representing a specific occurrence of a value situation that satisfies the value frame). Consequently, an occurrence of the `fs:Killing` frame satisfies the `fs:Killing` description, and in turn, being an occurrence of a Harm value frame situation, it satisfies a Harm description. By employing the punning technique in OWL 2, it becomes possible to treat the

[6] Prof. Keet online material can be found at https://eng.libretexts.org/Bookshelves/Computer_Science/Programming_and_Computation_Fundamentals/Book%3A_An_Introduction_to_Ontology_Engineering_(Keet)/07%3A_Top-Down_Ontology_Development/7.02%3A_Foundational_Ontologies

[7] DUL+DnS is available online here:https://ontopia-lode.agid.gov.it/lode/extract?url=http://ontologydesignpatterns.org/ont/dul/DUL.owl

same entity, identified by the same URI, both as a class (of situations) and an individual.

In this section, we have provided a brief overview of the primary semantic web resources utilized in this work, along with the commitment to the DOLCE foundational ontology. The subsequent section will focus on potential modeling pitfalls and challenges.

1.3 The Conceptual Modelling Bias Ladder

A fundamental prerequisite for Chapters 2, 3, and 4 is the recognition that when dealing with social sciences and humanities, especially when constructing domain ontologies that represent theories offering an incomplete or semi-formalized framework, the process of ontological transposition necessitates certain modeling choices.

It is important to note that we are not referring here to general modeling best practices, as discussed in Section 1.2.1, such as the utilization of Ontology Design Patterns [96, 143], leveraging existing well-known resources, or adhering to FAIR principles [120]. These principles should always be upheld by a good ontologist. The issue at hand is more nuanced and pertains to the multitude of challenges encountered when formally transposing a subject matter that may exhibit the following characteristics: (i) its formulation is obscure, as entire conflicting schools of thought can arise from different interpretations of one or more statements (e.g., the philosophical domain, where the contrasting positions of Hegelian Marxists and conservatives serve as a prominent example); (ii) the framework being considered may be internally incomplete in its axiomatization due to its expression in a non-formal language; (iii) it can also be internally inconsistent for the same reason mentioned above.

In addition to the aforementioned challenges intrinsic to the subject matter itself, we present in Figure 1.3 the instances of errors that directly depend on the ontologist. While not representing actual fallacies, it is crucial to acknowledge that each step illustrated in Figure 1.3 is indispensable for ontological transposition, and each step has the potential to introduce ambiguities and discrepancies. Therefore, each level of the Modelling Bias Ladder introduces an additional layer of bias. Consequently, the final product must be regarded as a translation, an approximation of the original, influenced by the modeller's unique understanding horizon [81].

The presented Conceptual Modelling Bias Ladder draws inspiration from above mentioned prof. Keet's online material.

The first step on the ladder is the foundational level, which involves making a crucial modeling decision. This decision entails committing to a foundational ontology as the upper-layer conceptual backbone, either explicitly (e.g., choosing between DOLCE and UFO [121]) or implicitly, such as excluding the representation of the temporal dimension. The foundational view typically includes the representation of the concept of parthood, which is depicted in Figure 1.3 as the Mereological issue.

The second step of the Modelling Bias Ladder addresses the aforementioned problems of conflict within a single theory or between sibling theories claiming to originate from the same ancestor.

Moving to the third step, we encounter the axiom level, where problems may arise

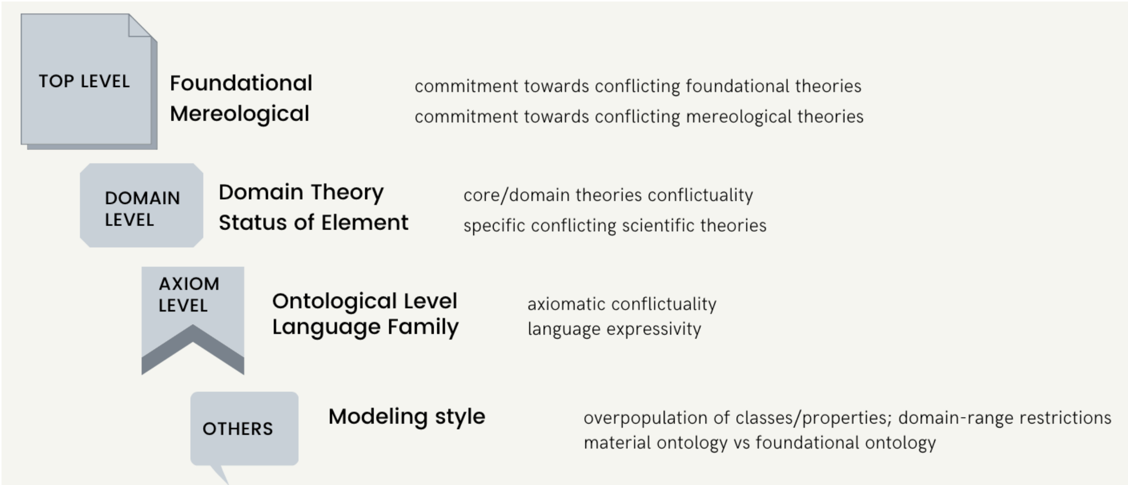


Figure 1.3: The Conceptual Modelling Bias Ladder

from the declaration of axioms that potentially introduce inconsistencies or from choosing one language over another. The choice of a more or less expressive language depends on the specific modeling purposes. For instance, in this particular work, the decision to use OWL 2 instead of a more expressive language like FOL is motivated by the aim to build an ecosystem and create integrated resources in the semantic web.

Finally, at the lowest step of the ladder, we have the modeling style, which reflects a preference for a verbose or essential approach, the level of class population, the degree of semanticization of properties, and the extent to which domains and ranges are restricted or left unspecified.

All of the aforementioned factors should be considered when interpreting the modeling presented in the subsequent chapters. To provide explicit examples of the modeling choices, we now present the Exuviae ontological module.

1.4 Exuviae Methodology for Epistemic Comparison

This section introduces an ontological module designed to formally represent concurring theoretical systems. Its purpose is not only to address the ontological question of ’what is what?’ but also to delve into the epistemological question: ’what is what *for whom*?’

The section focuses on an initial attempt to tackle this question within the ongoing and unresolved epistemological debate known as the “demarcation problem.” This problem encompasses the disagreement between philosophers of science and scientists regarding the distinction between what qualifies as science and what does not. By extension, it involves the differentiation of alternative theories that explain a similar set of phenomena.

While science, technology, engineering, and mathematics (STEM) disciplines prioritize causal explanations as the primary criterion for theory preference, as discussed by

Fleck [77], Kuhn [183], and Feyerabend [73], social sciences and humanities (SSH) often accommodate alternative theories that cannot easily be coerced into causal explanations. These theories pertain to phenomena such as emotional spectra, the motivation behind personal beliefs, moral foundations, historical or argumentative perspectives, and others. As outlined in the introduction of this work, these topics are prominent in the social information semiosphere prevalent in the so-called post-truth era.

The presence of alternative theories within SSH has created challenges in comparing them, particularly in the absence of causal criteria. Consequently, the concepts and semantics of relations within SSH are often unstable. Furthermore, without a causal foundation, semantic instability becomes a primary concern when comparing and selecting among alternative theories. In this context, the concepts and notions extracted from a reference system are referred to as "floating theory fragments." These fragments serve to categorize, explain, or even generate an empirical spectrum of phenomena.

Theoretical Implant and Context of Use Exuviae ("layers of skin or cuticle shed by animals during ecdysis"[8], or "In biology, exuviae are the remains of an exoskeleton and related structures that are left after ecdyso-zoans - including insects, crustaceans and arachnids - have moulted."[9]) is a computational ontology specifically designed to represent and elucidate the epistemic choices made during the modeling and comparison of elements from the same or different theories. It serves as a pragmatic logical framework, providing a conceptual exoskeleton for formally comparing sets of floating fragments.

The primary purpose of the Exuviae ontology is not to establish the superiority of one theory over another or to undermine theories developed in a less formal manner. On the contrary, it seeks to accomplish the opposite, aligning with Habermas' conception of social sciences and humanities (SSH). As Habermas states in his work [123]:

> Whereas the natural and the cultural or hermeneutic sciences are capable of living in mutually indifferent, albeit more hostile than peaceful coexistence, the social sciences must bear the tension of divergent approaches under one roof.

Exuviae allows us to construct an ontological container that facilitates the joint analysis of floating fragments, assesses their coherence, and potentially integrates them into a broader theory. It enables a comprehensive examination of diverse theoretical perspectives while maintaining a cohesive framework for the social sciences and humanities.

[8] https://www.collinsdictionary.com/dictionary/english/exuviae
[9] https://en.wikipedia.org/wiki/Exuviae

Methodology The Exuviae ontology serves as a conceptual exoskeleton and follows a methodology comprised of three main phases:

1. The initial step involves the explicit formal representation of a floating fragment. This entails modeling a theory, perspective, or interpretation using the OWL language. Concepts and relations within the fragment are identified and connected through logical axioms or vector spaces.

2. In the second phase, whenever possible, concepts and relations are aligned with foundational theories or specific reference domain frameworks (referred to as "core ontologies"). This alignment helps establish connections to ontological dimensions such as 2D or 3D entities, topology, mereology, identity, process models, participation, scalar models, common sense, or specific frames.

3. The final step encompasses formal comparison, resolution, and selection. The alignment achieved in the previous phase facilitates a formal comparison by providing correspondences as a backbone. Consequently, it becomes possible to determine the subject matter addressed by a fragment, including different types of entities, frames, and focal points. Similarities and clusters may emerge, along with potential equivalences, conflicts, and complementary aspects. This facilitates the integration of multiple floating fragments. Moreover, the criteria for the relevance of a fragment and its potential superiority over others can be identified, as well as any competition among fragments for the same role within a theory. Overlaps and explicit differences can also be addressed.

By following this methodology, the Exuviae ontology enables a systematic approach to analyzing, resolving, and selecting floating fragments, promoting integration and enhancing our understanding of complex theoretical systems.

Horizon of Understanding Exuviae is an ontology developed with the recognition that data constitutes a vast and complex domain, which may not lend itself to a singular, unequivocal interpretation. However, this acknowledgment should not lead to a simplistic relativism. The entire ontology, especially its object, data, and annotation properties, is designed to counteract a postmodern deconstructionist perspective that questions the framing of data. Exuviae strives to enable knowledge reasoning through a data-driven approach while explicitly documenting the reasoning process, assessments, methodologies, references, and evaluations involved in the epistemic comparison. The goal is to arrive at an interpretation as a justifiable understanding or, as advocated in philosophical hermeneutics, to "horizon of understanding" as articulated by Gadamer [81].

Additionally, Exuviae seeks to address phenomena such as “Conceptual Drift,” as discussed by Kuukkanen [185] and Wang [343], by representing specifications of extensional change in ontological form that correspond to the intensional modification of the concept itself.

Furthermore, Exuviae embraces Betti’s [23] perspective on conceptual modeling and interpretations and aims to operationalize it through its ontological structure:

> Making an interpretive framework explicit in fact provides the best defence against the risks of interpretative biases in the writing of intellectual history, and furthers the comprehension of texts.

By incorporating these principles and approaches, Exuviae strives to establish a robust foundation for knowledge representation and reasoning, mitigating biases, fostering understanding, and promoting a more rigorous and nuanced exploration of complex conceptual domains.

1.4.1 Ontology Structure

In this section, we provide a description of the Exuviae ontological module. This list of classes and properties is not intended to be exhaustive but rather open to the introduction of new entities, particularly properties, based on specific use cases.

Classes The main classes, depicted in Figure 1.4, are associated with and axiomatize the `:EpistemicComparison` class, which serves as the central component for comparison.

- `:TheoreticalFragment`: A TheoreticalFragment is a collection of ontology elements that are interconnected from one or multiple theories. The fragment serves as a central component for comparing different theories and deriving specific elements by applying epistemological choices based on certain criteria, potentially leading to a selection outcome.

- `:EpistemicComparison`: The EpistemicComparison class serves as the central component for conducting epistemic operations on theory fragments with the objective of comparing and selecting one fragment among others.

- `:CriterionMeasurement`: CriterionMeasurement represents the measurement of the criterion chosen to compare one Fragment against another. Examples of criteria include better literature grounding, increased soundness, better availability of resources, more operationalizable structure, etc.

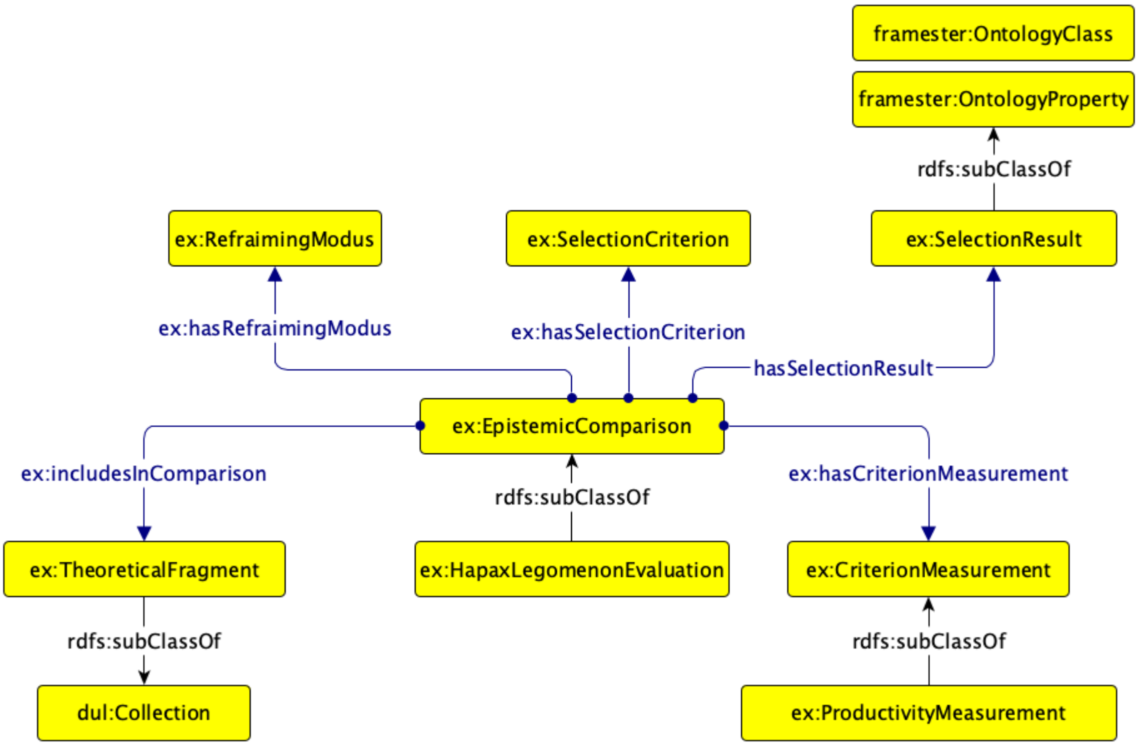


Figure 1.4: Exuviae Epistemic Comparison Hub.

- `:ReframingModus`: ReframingModus captures the way in which an entity is conceptually derived from a theory but with certain modifications.
- `:SelectionCriterion`: In cases where the ultimate purpose of employing the Exuviae ontology is the selection of a specific fragment from a set of alternatives, this class represents the criterion based on which the selection is made.
- `:SelectionResult`: After the selection process, this class represents the outcome of the selection based on a particular SelectionCriterion.

Object Properties The properties in Exuviae focus on the conceptual derivation of entities and aim to specify the type of conceptual dependence. The following properties are included:

- `:conceptuallyDerivedFrom`: This property is used to indicate the intellectual debt of one entity towards another. For example, a concept may have been developed based on a clue, intuition, concept, rule, or theory that provided input to a cognizer for the cognition of the concept. This property serves as a super-property for the following properties.

- `:contradicts`: This property indicates that some entity or part of it is derived and subsequently contradicted or negated, either partially or entirely.
- `:explicates`: This property is used to express a logical or formal conclusion that is explicitly stated by the conceptually derived entity but only implied, not explicitly stated, by the source.
- `:generalizes`: The derived entity corresponds to a broader extension than the source entity. This property is similar to the `skos:broadMatch` property, but specifically applies to conceptual objects.
- `:reframes`: This property signifies that the derived entity is conceptually derived from a source entity but reframed in some way. This could involve a stricter formalization, a different range or domain declaration, or any other form of reframing. If necessary, it is recommended to create subproperties that specify the type of reframing, if not already covered by other properties.
- `:reuses`: This property indicates that the derived entity is copied and reused (cloned), potentially with minimal nuanced distinctions, such as re-labeling without new semantic commitment.
- `:specifies`: The derived entity corresponds to a narrower extension than the source entity. This property is similar to the `skos:narrowMatch` property, but specifically applies to conceptual objects.

Data Properties The data property included in Exuviae is:

- `:explanation`: This property is used to provide a natural language explanation of a conceptual derivation declaration. Its purpose is to specify aspects such as the dimension of conceptual reframing, conflictuality, or any other relevant information that helps humans better understand the modeling choices. It is purely intended to enhance understanding for human users.

Annotation Properties The annotation property included in Exuviae is:

- `:bibRef`: This property is used to annotate the bibliographical reference of a fragment, concept, or even a whole theory. It includes information such as the original definition, detailed occurrences reference location record, and the year of publication. The purpose of this property is to track back the original sources and allow users to understand and retrieve the original information related to each modeled entity.

Exuviae Application In this section, we have introduced Exuviae, an ontological module designed for epistemological comparison, enabling the ontological representation of theories and facilitating formal conceptual transfer between modules or theories.

The process of transposing theories into an ontological framework, which is a preliminary step for conducting conceptual analyses using the Exuviae exoskeleton, should adhere to good modeling practices. These practices have been discussed in Section 1.2 and are outlined in works such as [92], [62], [66], [268], and [350].

An analogy to an old Persian tale [2] can help illustrate the role of Exuviae's ontological module. In the tale, three brothers and 17 camels face the challenge of dividing the herd according to a specific rationale. However, a mathematical division of 17 camels among three brothers is impossible. To resolve the issue, a wise man introduces an 18th camel, enabling the distribution of the herd among the brothers. Similarly, the Exuviae ontology serves as the 18th camel, providing a solution to the problem of floating fragmentation in social sciences and humanities. It acts as an exoskeleton for the floating fragments, allowing us to approach the phenomena from a different perspective and potentially integrate or select the best fragments for disciplinary advancement or specific tasks.

In summary, Exuviae is an ontological module that facilitates formal epistemic comparison. It plays a crucial role in the creation of the ISAAC ontology, as discussed in Chapter 2, where diverse theoretical contributions are harmonized into a single module by aligning various contributions from different sources.

The next section will present the QUOKKA workflow, which involves populating knowledge graphs by reusing entities from semantic web resources integrated in the Framester hub.

1.5 QUOKKA workflow

In this section, we introduce the QUOKKA workflow, which is a semi-automatic workflow designed for QUerying Ontological resources and Knowledge bases for Knowledge Augmentation. The workflow involves a series of SPARQL queries aimed at query expansion, with the objective of constructing a semantic frame within a desired domain.

The main goal of the QUOKKA workflow is to populate knowledge graphs with Linked Open Data entities sourced from reputable and high-quality semantic web resources. This process enables the creation of versatile and comprehensive knowledge graphs, incorporating thousands of semantic web triggers relevant to any given domain.

The conceptual inspiration behind the QUOKKA workflow is derived from the notion of *Qualia structure*. The term "Qualia" originates from Aristotelian categories but has been redefined within the context of cognitive sciences and neurosciences. In this regard, we refer to Pustejovsky [260] and, specifically, his classification of four main types of Qualia:

1. *Formal Qualia*: This refers to the ontological status of a semantic domain and encompasses the portion of reality covered by that domain.

2. *Constitutive Qualia*: This encompasses the mereological and partonomic relationships associated with a semantic domain.

3. *Telic Qualia*: This perspective focuses on the teleological aspects and aims associated with an entity within a semantic domain.

4. *Agentive Qualia*: This pertains to the entities that participate in the constitution of an entity. In the case of a conceptual frame abstraction, it refers to the semantic role structure involved in the occurrence of a frame situation.

By drawing inspiration from these Qualia types, the QUOKKA workflow enhances the process of knowledge augmentation and semantic frame construction within knowledge graphs.

The resulting workflow, depicted in Figure 1.5, serves the purpose of populating knowledge graphs in the main domains addressed in this work, namely image schemas (Chapter 2), values (Chapter 3), and emotions (Chapter 4).

This section specifically focuses on explaining the underlying rationale employed in populating the knowledge graphs. In this work, all the modules that operationalize a particular theory are populated with entities obtained from semantic web resources integrated within Framester. Therefore, the SPARQL queries conducted, which are available in Appendix **??** and on the QUOKKA GitHub repository[10], adhere to the semantics used in

[10]The QUOKKA GitHub is available here: `https://github.com/StenDoipanni/QUOKKA`

the Framester resource and generate results in accordance with it. However, the overarching objective of the QUOKKA workflow is to establish a more general frame building workflow. This workflow aims to address the competency questions that arise during the process of gathering knowledge within a specific domain and organizing it within a frame structure, particularly by leveraging pre-existing resources.

By following the QUOKKA workflow, users can systematically gather and organize knowledge about a given domain, while reusing and integrating existing resources into the frame structure.

1.5.1 Frame Building Workflow

Following the structure depicted in Figure 1.5, we present a list of semantic relations that can be explored to determine the relationship between an entity and a domain. Each paragraph corresponds to a SPARQL query executed at the Framester endpoint, aimed at retrieving the entity type indicated in the paragraph title. These queries are designed to identify semantic triggers associated with each entity type. All the queries are available and documented on the QUOKKA GitHub repository.[11]

Manual Lexical Units Selection The initial step in the process involves a manual selection of a limited set of lexical units that are directly related to the conceptual frame being constructed. It is recommended, following the XD methodology outlined in Section 1.2.1, to engage domain experts in identifying terms that precisely capture the targeted portion of reality to be encompassed by the frame. This set of lexical units is referred to as the Starting Lexical Material (SLM) set. Detailed information about the specific SLM set and potential expansions for each domain of interest in this study can be found in Chapters 2, 3, and 4.

The SLM set is further expanded using the WordNet resource, which can be accessed through its online user interface[12]. This expansion aims to leverage the relationships of hyponymy and synonymy among terms. The rationale behind this approach is that if a particular domain frame is evoked by a given lexical unit, it logically follows that it should also be evoked by more specific terms (though the applicability of this principle may depend on the specific frame and domain being modeled). The domain of image schemas offers numerous examples of this phenomenon. For instance, if we consider the schema used to conceptualize "movement" from a frame semantics perspective, and we accept that it is evoked by the term "walking," then we must also accept that it is evoked by terms such as "running" and "pacing" as well.

[11] The QUOKKA GitHub is available here: `https://github.com/StenDoipanni/QUOKKA`

[12] WordNet online resource is available here: `https://wordnet-rdf.princeton.edu/`

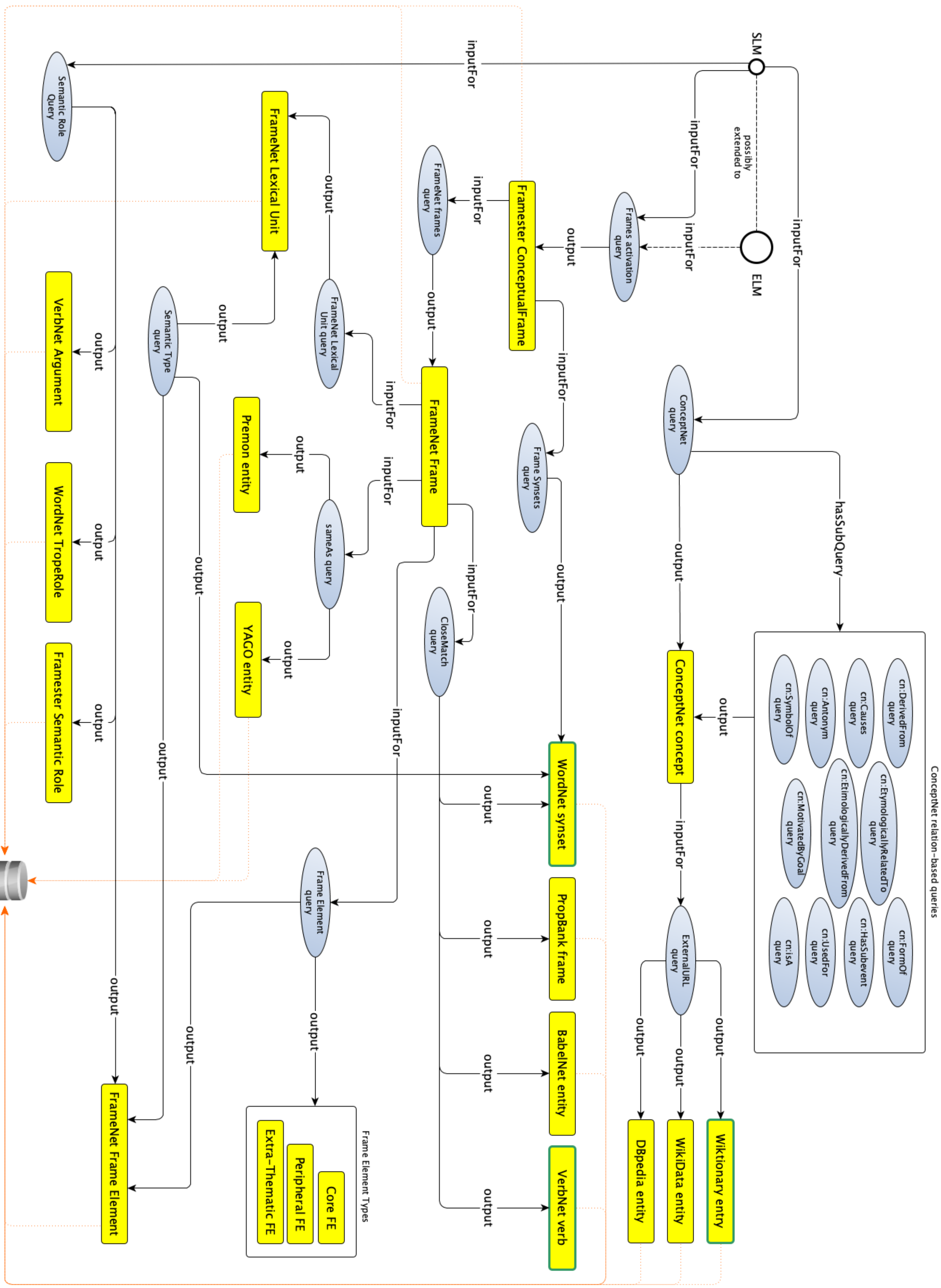


Figure 1.5: General QUOKKA frame building workflow.

By manually selecting an initial set of lexical units and expanding it using WordNet, we can lay the foundation for building a robust conceptual frame for the desired domain. The specific SLM set and the results of the expansion process play a crucial role in this endeavor.

ConceptNet-driven triggering The next step involves utilizing ConceptNet relations using each entry from the SLM set as an input variable. The current version of ConceptNet in Framester is ConceptNet 5, while version 5.5 is available on the ConceptNet online resource[13]. The following semantic relations defined in ConceptNet are reused in this query:

- `cn:DerivedFrom`: Indicates that a concept is derived from another concept.
- `cn:Causes`: Specifies that an entity is the cause of a concept.
- `cn:Antonym`: Expresses that an entity represents a polarity opposite to the one described by a concept.
- `cn:isA`: Indicates that an entity is subsumed by a concept.
- `cn:EtymologicallyRelatedTo`: Establishes an etymological relationship between the lexical unit referring to an entity and the one referring to a concept.
- `cn:SymbolOf`: Indicates that an entity serves as the symbol of a concept. In Peirce's terms, it is more accurate to say that an entity is the *icon* for a concept, as this relation is primarily used to connect concepts with emojis representing them.
- `cn:UsedFor`: Specifies that an entity has functional or instrumental relevance to a concept.
- `cn:HasSubevent`: Identifies that an entity, considered in its temporal extension as an event, is a sub-event of a concept.
- `cn:MotivatedByGoal`: Indicates that an entity is motivated by a concept, conceptualized as the ultimate goal, for which some form of intermediate step is necessary.
- `cn:EtymologicallyDerivedFrom`: Highlights that a lexical pointer is etymologically derived from another lexical material that refers to a concept.
- `cn:FormOf`: States that an entity is a variation or form of a concept, implying a comparison process that identifies differences on physical or non-physical dimensions.

[13] ConceptNet can be accessed at `https://conceptnet.io/`

By leveraging these ConceptNet relations, we can explore the interconnectedness between lexical units and concepts, expanding the initial set of terms and enriching the conceptual frame construction process.

Furthermore, these properties enable the exploration of the ConceptNet network of relations and facilitate the extraction of entities that are aligned with other widely recognized semantic web and multi-modal resources. As a result, the initial Starting Lexical Material (SLM) set can be expanded to a much broader triggering set, encompassing a wider range of related entities. This expansion enhances the comprehensiveness and richness of the conceptual frame construction process.

WikiData lexical triggering The WikiData lexical triggering step relies on the entities obtained from the ConceptNet-driven triggering process. The Starting Lexical Material (SLM) set serves as the input variable to retrieve all corresponding WikiData entries. This step enhances the integration of lexical information from the WikiData resource, contributing to the enrichment of the domain knowledge base.

DBpedia factual triggering In parallel with the WikiData lexical triggering step, it is possible to retrieve entities aligned with the DBpedia resource, which provides factual grounding for the domain knowledge base. Starting from ConceptNet concepts, this process allows for the identification of relevant entities in the DBpedia resource. The query responsible for this step is illustrated in Figure 1.5 as the “DBpedia External URL query”.

Frame-driven triggering In a separate branch, as depicted in Figure 1.5, the selected lexical units from the Starting Lexical Material (SLM) set are examined to determine if they serve as lexical triggers for existing FrameNet frames. This step aims to leverage pre-existing frames that may partially overlap with the desired domain or provide more specific or general situational schematization.

In the Framester resource, the lexical units, without disambiguation, are used as variables to collect all their senses and the frames evoked by each sense. The query responsible for this step is represented by the “Frames activation query” starting node in Figure 1.5. The number of senses and related frames may vary depending on the input lexical material. To improve data quality and distinguish relevant semantic senses from unrelated ones, further contextual disambiguation is required.

After executing the SPARQL query, the set of frames selected as potential triggers is manually determined. As mentioned in the paragraph on Manual Lexical Units selection, once the query iteration is repeated for all synonyms and hyponyms, the initial step of the QUOKKA frame building workflow can be considered complete, allowing for progression to the subsequent step.

Frame element-driven triggering In the event that the domain to be modeled encompasses a broad scope, certain aspects of it may already be addressed by existing frames, enabling the adoption of the established structure of formalized semantic roles. This step pertains to such a scenario, where frame element activation revolves around the activation of semantic roles associated with the occurrence of a `dul:Situation`, i.e., a Frame occurrence, within the specified domain.

The key conceptual inquiries are consequently concerned with the structure of semantic relations implicated in the domain, such as identifying the necessary roles or determining the presence of an Agent/Patient, among others. The corresponding SPARQL query is focused on retrieving FrameNet frame elements of types "Core," "Extra-Thematic," and "Peripheral." These encompass element types that are ontologically essential to the frame occurrence, those that specify crucial characteristics pertaining to the situation or the roles involved (e.g., action degree, intensity), and finally, roles that typically participate in frame occurrences, including temporal and spatial aspects.

In this work, the structure of semantic roles is inherited from both FrameNet frame elements and their alignments with the Framester resource. The query associated with this step is represented by the "Frame Element Type Query" in Figure 1.5.

FrameNet Lexical Units triggering This step naturally follows from the preceding paragraphs and the adoption of a frame structure. By accessing the online user interface of the FrameNet resource[14], it becomes apparent that certain lexical units are identified as evoking specific frames. To fully comprehend the semantics of these lexical units, a common understanding of the system of semantic relations and the contexts in which they convey their meanings is required.

The purpose of this SPARQL query is to incorporate the FrameNet lexical units as triggers for the frame being constructed. These lexical units are declared in FrameNet as triggers for frames that exhibit either total or partial overlap with the intended domain. (The extent of overlap determines the necessity of involving human experts in order to ensure data quality.)

The specific SPARQL query, referred to as the "FrameNet Lexical Unit Query," is presented in Figure 1.5. It proves beneficial for resources that align with FrameNet and employ lexical unit retrieval for frame detection purposes.

WordNet lexical triggering The activation of lexical material plays a significant role in semantic detection, and it is accomplished by automatically reintroducing the results of the Frame activation query into the workflow. This refers to the frames that were previously

[14]The FrameNet online user interface can be found at `https://framenet.icsi.berkeley.edu/fndrupal/about`

manually selected. The rationale behind this step is that if an entity evokes a FrameNet frame that is related to the frame being modeled, then that entity should also be activated in relation to the frame itself.

Therefore, this query expands the lexical coverage beyond FrameNet lexical units by including other well-known semantic web resources. It retrieves all the elements that evoke a frame. In the Framester semantics, all WordNet synsets are considered frames as well. A synset represents the class of situations to which a specific sense of a term is applicable. This alignment enables the clustering of lexical units with a particular sense, which in turn evoke a specific frame. The synset, representing the class of situations associated with a given meaning, is subsumed by the frame, representing the class of situations satisfied by the occurrence of the frame. The SPARQL query aims to broaden the lexical coverage for all the senses of a specific set of terms, considering their contextual usage.

It is important to note that the number of elements retrieved can be substantial, especially for broader frames. The set of triggers can potentially include thousands of WordNet synsets. Synsets are extensively used in various works, including large multimodal resources [181], for tasks such as disambiguation, alignment, and entity recognition. They constitute a vital component of the knowledge base to ensure comprehensive coverage and facilitate the proper operationalization of the frame.

The WordNet version aligned with Framester is version 3.0, while the version available from the WordNet repository is 3.1. It is worth mentioning that there may be differences in coverage and extension between the two versions.

The specific SPARQL query, referred to as the "Frame Synsets Query," can be found in Figure 1.5.

Close Match triggering In addition to WordNet synsets, entities from various semantic web resources are aligned with frames in the Framester hub. This alignment is established at the meta-level using the `skos:closeMatch` object property. It declares that a concept identified by a specific URI in one resource has a close match with another concept identified by a different URI in another resource. Although the two entities remain distinct, they point to the same or similar aspect of reality.

Here, we specify the entities aligned with the `skos:closeMatch` relation from several resources, how they interlink with each other, and the specific SPARQL query for each resource:

- *WordNet synsets*: These are sets of contextual synonyms. As explained in the previous paragraph, if two lexical units can be used as synonyms in the same context, it can be inferred that the considered context is possibly a subframe of the frame being modeled. Therefore, declaring the entire synset as a trigger for a frame results in a significant increase in coverage, including all the senses of the terms that can be used in similar situations. Some frames that schematize events or actions may have a `skos:closeMatch` relation to verbs or nouns that point to those events or actions. The query to retrieve WordNet synsets subsumed by a frame is the one mentioned in the previous paragraph, while the general query to retrieve those aligned via `skos:closeMatch` is mentioned at the end of this paragraph.

- *VerbNet verbs*: Verbs from the VerbNet resource can be retrieved through the alignment between WordNet "word senses" and VerbNet "key senses", as well as through the close match alignment with frames. The query to retrieve VerbNet verbs is shown in Figure 1.5 as "VerbNet triggering."

- *PropBank frames*: Frames from the PropBank resource are aligned with FrameNet frames through the `skos` relation. By providing the URIs of the FrameNet frames as input for the "PropBank triggering" SPARQL query, entities from the PropBank resource can be collected.

- *BabelNet entities*: A further multilingual coverage is provided through the alignment of BabelNet with Framester frames. The updated online version of BabelNet (5.2) may differ in size and coverage compared to the version in the Framester resource (3.7). Nonetheless, it is possible to retrieve entities from over 270 languages through the `skos:closeMatch` alignment.

- *Premon entries*: Premon entries are an extension of the lemon model by the W3C Ontology - Lexica Community Group.

Entities from all the mentioned resources can be retrieved using dedicated queries or through the `skos:closeMatch` query, depicted as a blue oval labeled "CloseMatch Query" in Figure 1.5.

YAGO Ontology triggering The WordNet lexical grounding is utilized once again to retrieve entities from the YAGO (Yet Another Great Ontology) resource. In this case, the alignment is achieved through the `owl:sameAs` property towards WordNet synsets. The corresponding query is presented in Figure 1.5 as "YAGO Ontology query."

Semantic role-driven triggering While potential roles participating in a specific frame are extracted from FrameNet frame elements (as described in the "Frame element-driven triggering" paragraph), according to Framester semantics, they are not the only sources for structural elements that can serve as roles in a frame occurrence.

In Framester, triggering assertions from FrameNet frame elements are extended to include multiple sources of semantic roles: VerbNet arguments, PropBank roles, and WordNet tropes. Semantic roles in Framester are organized within a complex taxonomy, with a small top-level that aids in integrating and activating them.

To retrieve the semantic roles participating in a frame, two queries are executed, starting from the top nodes of different graphs. This approach ensures the activation of both general and specific roles.

These queries can be performed by either starting from manually selected frames and inheriting the semantic roles or starting from the SLM (Selectional Linking Model) used as the input variable to directly retrieve roles independently of the frame of origin. It is possible that the to-be-modeled frame is evoked only by a certain role of an already existing frame (e.g., frames representing two plausible opposite outcomes of a particular situation). Therefore, human involvement is necessary during this step of populating the knowledge base. The potential outcomes of these queries are displayed in Figure 1.5 as the output of the "Semantic Role query."

Semantic type-driven triggering A dimension that complements the previously mentioned aspects is the semantic type of an entity. Although semantic types are more applicable to physical dimensions (e.g., spatial types like `fnst:Front`, `fnst:Back`) or sensorimotor aspects of entities (e.g., `fnst:Source`, `fnst:Goal`), they should be considered in the overall frame building workflow when populating the knowledge graph to operationalize the frame.

This step involves retrieving all existing FrameNet semantic types and manually exploring their differences and coverage to select the semantic types that may trigger the desired frame. Subsequently, a second query is executed to find entities filtered by the iteration of non-disambiguated lexical units from synsets and their hyponyms. This query also extracts the semantic type of the entities. Finally, a coherence check is performed between the retrieved entities, their semantic type, and their evocation of the frame.

The queries for this step are displayed in Figure 1.5 as "Semantic Type query." This final query, which relates to Pustejovsky's Type theory [259], is particularly significant. It is used to infer additional knowledge from graph pattern inferences related to *Type Matching*, *Type Accommodation*, and *Type Coercion*, as explained in Chapter 5.

A necessary concluding note regarding the QUOKKA workflow is that when dealing with a significant amount of resources, which includes thousands of triples, and especially when dealing with conceptual matters, it is important to consider the potential noise that can arise due to various factors such as different levels of accuracy, multiple alignment steps, misunderstandings, and diverse purposes.

To mitigate the introduction of excessive noise, a graph accuracy scoring system was developed in this work. It is integrated into the Framester resource as a scale ranging from 1 to 5, and it is represented in the ontology through the data property `fschema:reliabilityScore`. When applying the QUOKKA workflow to the Framester resource, it is advisable to set the value of the mentioned data property as >= 4 for all the queries. This setting helps to establish a filtering granularity that maximizes both accuracy and recall.

It is worth noting that all the queries presented in Figure 1.5 already incorporate the reliability filter in their online version available on the QUOKKA GitHub.
The subsequent section is dedicated to ontology testing methodologies.

1.6 Ontology Testing

In this section, we provide information regarding the inference testing conducted for each module described in the subsequent chapters. Additionally, we describe the process of generating a knowledge graph from natural language and extracting knowledge related to image schemas, values, and emotions.

1.6.1 Automatic Evaluation Method

To test the resources proposed in Chapters 2, 3, and 4, we employed an automatic extraction approach for image schemas, values, and emotions. This process involved generating knowledge graphs automatically from natural language and performing automatic detection through entity linking, frame detection, and word-sense disambiguation.

Research in generating graphs from textual data is a well-explored area, with various frameworks and approaches available. Some frameworks, such as those proposed by Petroni et al. [243], Roberts et al. [269], and Shin et al. [302], utilize large language models to query and extract knowledge bases comprising factual and commonsense knowledge. Wang et al. [342] proposed a method involving attention weight matrices, while Melnyk et al. [221] recently tested a new architecture using the Amazon AI model as a baseline [122], yielding promising results.

However, for the purpose of this work, it was not only necessary to automatically generate knowledge graphs from text, but also retrieve entities from the semantic web resources mentioned in Section 1.5. For this purpose, we utilized the FRED tool [99] in our resource testing.

FRED can be described as a "situation analyzer" that employs a combination of statistical and rule-based components to extract knowledge from natural language. It generates RDF/OWL knowledge graphs and incorporates entity linking, word-sense disambiguation, and frame/semantic role detection. Since FRED is directly linked to the Framester ontology, its graphs include the following features: (i) word sense disambiguation using the WordNet resource, (ii) disambiguation of VerbNet verbs, including the assignment of VerbNet semantic roles to the sentence's semantic argument structure, (iii) frame detection from FrameNet, (iv) recognition of PropBank frames, and (v) entity linking to DBpedia.

To illustrate, let's consider the sentence: 'Italian politics risks being overwhelmed by the new wave of disgusting corruption.' The knowledge graph automatically generated by the FRED tool for this sentence is shown in Figure 1.6.

In the graph, the root node represents the main verb "to risk," which, following the intensionality and extensionality semantics discussed in Section 1.2.2, corresponds to a local occurrence of the "Risk" verb. It is disambiguated to the VerbNet entity `vn:Risk_94000100`.

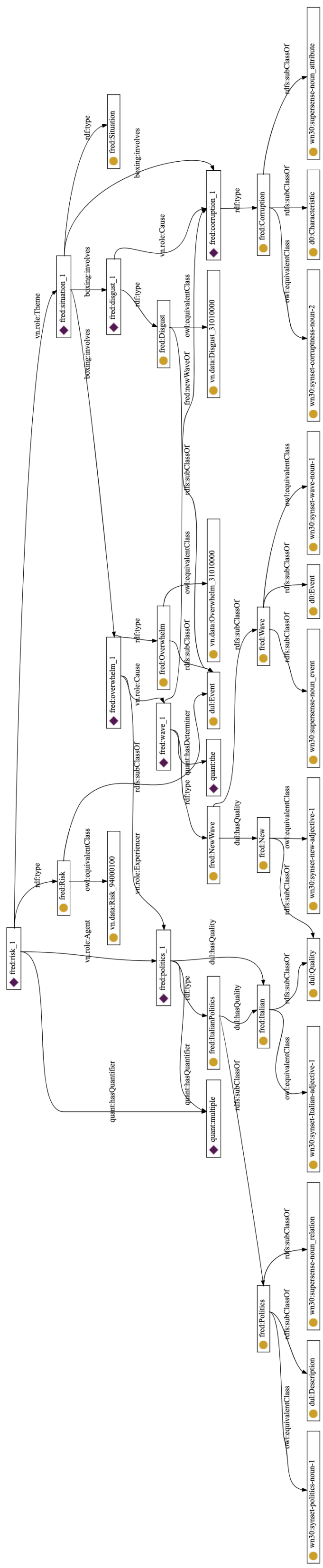


Figure 1.6: FRED graph automatically generated for the sentence: *Italian politics risks being overwhelmed by the new wave of disgusting corruption*

Following the paths from the root node in the graph, we see that the VerbNet role Agent is filled by the entity "politics," represented as the subject of the verb "risks." It is disambiguated to the WordNet synset `wn:politics-noun-1`. On the right side of the graph, the VerbNet role Theme is filled by a situation (an occurrence of an unspecified frame) that involves "disgust," disambiguated to the VerbNet entity `vn:Disgust_31010000`. This "disgust" is caused by "corruption," recognized as the type `wn:corruptness-noun-2` in WordNet. Finally, the "politics" serve as the Experiencer of an "overwhelm" event, where the cause of this event is a "wave" represented by the WordNet synset `wn:wave-noun-1`. We will leverage the formal semantic structures, similar to the one illustrated in Figure 1.6, to test our resources by incorporating image-schematic, value, and emotion layers on top of the existing knowledge.

Each detector (image-schematic, value, and emotion) follows the same pipeline, consisting of three main steps:

1. The detector takes a sentence in natural language as input, which is then passed to the FRED tool. FRED parses the sentence, builds a knowledge graph of semantic dependencies based on its syntax, and performs various tasks such as frame extraction, WordNet disambiguation, and entity recognition.

2. In the second step, the detector navigates the graph generated by FRED and focuses only on the entities retrieved from reused resources such as FrameNet, WordNet, VerbNet, and DBpedia. It disregards other nodes and arcs in the graph. For each of these relevant nodes, the Value Detector performs a SPARQL query to the MFTriggers graph to check if there are any semantic triggers associated with certain values.

3. Finally, for each successfully retrieved triggering occurrence, a triple is added to the original graph to indicate the triggering. The output is an "kg.ttl" file containing the graph generated by FRED and all the localized instances of value activation.

Detailed results and statistics regarding activation and triggering are provided in each chapter for each resource. An online version of the image schematic, value, and emotion detector is available, and they can be used separately or in combination to intertwine multiple layers of knowledge, enhancing the inferential power of SPARQL queries.

1.7 Chapter Conclusions

This chapter has provided the necessary background information to comprehend the key theoretical perspectives employed in this work, the addressed topic, and the investigated domains. It has also offered technical insights into the concept of ontology, the structure of the Framester ontology, the utilization of various semantic web resources, the DOLCE foundational ontology, and the XD methodology employed to develop ontological modules and populate knowledge graphs adhering to sound modeling practices.

Furthermore, we introduced the FRED tool and the graph-based pipeline for value detection, which is used for knowledge extraction from natural language and for the phase of ontology coverage testing.

In the subsequent chapter, our focus will be directed towards the first of the three primary domains explored in this work, namely the domain of sensori-motor cognitive patterns. Specifically, the next chapter will delve into the details of the Image Schemas Abstraction And Cognition (ISAAC) ontology.

Chapter 2

Image Schemas: the Basic Building Blocks of Cognition

This Chapter is dedicated to embodied cognition and in particular to Image Schemas. In Section 2.1 several works are mentioned, including definitions and description of the main focus and differences case by case; the final paragraphs are dedicated to image schemas formalisation and transposition in axiomatic language, as well as experiments and methodologies of detection from text. In Section 2.2 the ISAAC (Image Schema Abstraction and Cognition) ontology is presented. Note that, being some modules of the ISAAC ontology a transposition in ontological form of some of the main theories, the detailed description of aspects related to specific image schemas, their mereological aspects, the relations possible among them and clarifications about their ontological status is demanded to the specific section dedicated to the ontological module formalising the theoretical fragment in exam.

2.1 Image Schemas Theoretical Grounding

Embodied cognition is a *florilegium* of theories with several major currents, but all of them start from the pivotal idea that our thoughts and understanding of the world are determined by how we bodily experience and interact with the environment. It was originally developed by Mark Johnson and George Lakoff [159, 190, 191] as well as Varela's "Embodied Mind" assumption [333], and it has been strongly backed by last decades of scientific research, thanks also to fMRI experiments and neuroscience research [83]. The connection between language and factual knowledge of the world has been largely investigated in several research areas such as neuroscience [326, 101, 3], cognitive linguistics [86, 84], philosophy [300] and developmental psychology [213].

Embodied cognition theory suggests that our understanding of physical dynamic structures, namely, our introjection of perceptual information and learning from repeated experience of physical dynamics, ground our way of conceiving abstract concepts, such as time, ideas, emotions, morality, etc., in particular this structured information transfer has been studied in *Conceptual Metaphors Theory* [188, 189]. Conceptual metaphors are, in fact, metaphors based on (and developed as a consequence of) our bodily experience, functional domain mappings that allow predicating something about an entity from a domain using terms from another. The literal truth value of the statement, as in all metaphors, is false, but we are able to understand the meaning of the expression thanks to these powerful versatile schematic structures. For example, the conceptual metaphor `IMMORALITY IS A DISEASE` allows to express well formed statements such as 'This behaviour is sick and disgusting!', predicating something about an immaterial entity (behaviour), and performing an appraisal on it, declaring some qualities that define it as a manifestation of a disease while triggering some disgust. Resources of conceptual metaphors are available [48, 63], and have been formalised as ontological modules [97]. Another famous example among conceptual metaphors related instead to the spatial dimension is `TIME IS A LANDSCAPE THROUGH WHICH THE OBSERVER MOVES`, which maps time, not perceivable directly by our sensori-motor system, to a dimension that our entire body is well prepared to perceive: space. We can therefore speak about *travelling through* time, *going back and forward* and having time *loops*.

Ultimately, the embodied cognition theory has helped to shift the focus from internal, brain-based processes, totally untouched by the external reality, to the role of our bodily experiences and interactions with the environment, and their being determinant in shaping thoughts and understanding. Since the aim of this work is to realise some cognitively inspired artificial intelligence system to investigate the moral/cultural latent commonsense knowledge in techno-social environments, it seems proper to start from embodied cognition theory, and in particular to focus on that part of the theory, which aims at providing an explanation to the (still open) problem of the connection between embodied experience and cognitive representation.

The term "Image schema" has been used for the first time by Lakoff and Johnson in their *The Body in the Mind* [159] and elaborated in *Philosophy in the Flesh* [191]. It is inspired by the "Kantian schemata", from Immanuel Kant [165], who, in his *Critique of Pure Reason*, introduces them as bridges between the world of pure, non-empirical categories, and the human, individual, sensual experience. The title of this Chapter is taken instead from the definition that Jean Mandler gives to Image schemas as "the first conceptual building blocks" [212]. Image schemas have been proposed within the tradition of embodied cognition as conceptual structures that capture sensori-motor experiences and shape abstract cognition, including commonsense reasoning and semantic structures of natural language

(see e.g. [214, 321]).
They are in fact defined by Johnson as:

> recurring, dynamic pattern of our perceptual interactions and motor programs that gives coherence and structure to our experience.

while the "image" in "image schemas" is inspired by Talmy's works [320], [319] on spatial and force relations in language, referring to "images" as conceptual primitives. Although previous works have proposed an IS inventory [39, 215], a generally agreed upon final list is still open to debate and Johnson's [159] caveat still stands:

> What is important is that these recurrent patterns are relatively few in number, they are not propositional in the Objectivist sense, and yet they have sufficient internal structure to generate entailments and to constrain inferences (and thus to be propositionally elaborated).

Image schemas are said to be internally structured *gestalts* [336], that is, composed by spatial primitives (SP), introduced thanks to Mandler [215] that make up more complex image schemas as unified wholes of meaning [131, 214, 138]. To provide some example: a "bottle" is a Container with an inside, an outside, and a border containing liquids in the physical world. By way of metaphorical projection these characteristics are captured by the image schema Containment that is transferred to the abstract realm of emotions inside a body and linguistically expressed, e.g. to *bottle up*, meaning not making emotions leave some Container, realising the cognitive metaphor `BODY IS A CONTAINER FOR EMOTIONS`. While their existence in natural language has been studied by means of corpus-based (e.g. [240, 255]) and machine learning methods (e.g. [115, 116, 339]), few approaches to formalize image schemas (e.g. Image Schema Logic [137]) and connect them to existing resources to capture semantics exist. Previous work on image schemas and ontologies focuses on formalizing specific IS, e.g. Containment [136], or a specific perspective, e.g. IS as families of micro-theories [138], where authors exemplify their perspective or theory based on (possible combinations of) specific image schemas.

In terms of dynamic aspects, Galton [87] and Steedman [311] investigate the notion of *affordance* in relation to image schemas. Affordances as defined by Gibson [105] concern commonsense about the opportunities for action offered by real world objects, environments and roles. Schorlemmer et al. [285] investigate the characterisation of creative processes in conceptual blending [72] by means of diagrams of image schemas. A diagram is intended here within the context of category theory, and a means to model the internal structure of a categorical object. Such framing of image-schematic diagrams within a category-theoretic model of creative processes seeks to provide a mathematically rigorous and computationally feasible model of image-schematic structures.

The work by Kuhn [184] is relevant in using WordNet to extract the image schematic structure from expressions and concepts, followed by formally representing the extracted image schemas using the Haskell programming language. Walton and Worboys [340] advance on this work by aiming to express how image schemas are not only connected to one another, but can be combined to form complex conceptualisations. Further works related to affordance-related aspects of meaning inferred thanks to image schematic knowledge and applied to the spatial dimension in architecture can be found in Borgo, Melone and Kutz [222].

Several approaches have used image schemas to model events and scenarios (e.g. [309, 22]) starting from compositionality of IS, like Object, Path and Contact to obtain more complex ones like Blockage, Bouncing and Blocked_Movement, introducing temporal dimension. Other formal work includes how to structure IS as families or clusters of similar concepts (e.g. [138, 281]). Bennett and Cialone [17] elaborate on this idea of clustering and analyze occurrences of Containment in biological textbooks in order to propose method based on sense clusters for semi-automatic construction of a spatial ontology from natural language. It focuses on senses as contextualized interpretations, and expresses them within RCC-8 [264] to formally represent different spatial configurations of spatial primitives within the context of Containment.
Kimmel's work [173], extending Gibbs [103], adopts a sociological perspective and introduces the notion of IS as being socio-culturally situated. The idea of being able to understand image schemas only considering them *situated* in a socio-cultural environment is in line with the main idea of this work, which, borrowing terms from logic, referring to Chapter **??**, considers image schemas universal in their intension, but socio-culturally situated in their extension; namely they exist as hard-wired structures and evolutionary heuristics, but their realisation and consequent expression in language depends on the socio-cultural conceptualisation of the domains involved in the occurrence of usage.
Therefore, to summarize general pillars about image schemas from literature since here mentioned:

1. Image schemas are sensori-motor cognitive patterns that shape our perception and conception of the world;
2. They are embodied, meaning grounded in, and depending on, our bodily perception and shape;
3. They are "dynamic", meaning that they can change over time in response to new experiences and learning;
4. They are *gestalt* entities, meaning they have parts and components, but they cannot be conceived but as wholes of meaning;

5. They are hierarchical, meaning that they can be combined and structured into more complex patterns.

IS Formalisation Albeit IS seem to be recognised as pivotal for grounding the relation between external world perception, and human conception, by mean of not yet fully clarified analogy-based cognitive process, very few attempts to formalise the IS domain can be found. To provide a more formal account, Hedblom et al. [138] propose the unified metalanguage Distributed Ontology, Modeling and Specification Language (DOL) [229] to represent shared gestalt structures of seemingly unrelated image schemas as a family, that is, a set of interlinked theories. Such a gestalt grouping of experiential structures implies a distinction between primitive and complex types. To this end, the theory of Mandler and Pagán Cánovas [215], rooted in developmental psychology, was adopted to distinguish between spatial primitives, image schemas, and conceptual integrations. Spatial primitives are the very first preverbal building blocks that infants form to quickly compose more complex structures, i.e., spatial primitives are the parts that compose coherent unified wholes. These wholes (or spatial events) built from spatial primitives are image schemas. Finally, conceptual integration refers to the inclusion of non-spatial elements, such as values, or emotions. Hedblom et al. [138] take up this initial definition and depict spatial primitives as roles participating in a frame's image schema (e.g. the spatial primitives , Path and Goal are roles of the Source_Path_Goal image schema).

Thereby, image schematic-structures, primitive or not, can be grouped based on experiential gestalt family resemblances. Hedblom formalizes also IS combinations dynamics, in particular: (i) "Collection": a set of IS which do not alter the gestaltic properties of a particular spatio-temporal relationship, but together are able to represent a particular experiential structure; (ii) "Structured Combination": similar to a Collection, but with the relevant addition of a sequential cause-effect relation; (iii) "Merge": the combination of IS in such a way that gestaltic properties are altered [139].

This initial formalisation is later extended as the formal language Image Schema Logic (ISL^M) [137]. It brings together RCC-8 [264], Qualitative Trajectory Calculus [345], cardinal direction, and linear temporal logic [182, 266]. ISL^M is exemplified with the formalization of Support and Contact, and later applied to Containment [136] as well. While ISL^M provides strong formalization for (combinations of) image schemas, nevertheless, a direct link to natural language and proper lexical coverage is still missing.

The following presents a list of IS, for which ImageSchemaNet currently provides both a formalization and a lexical coverage, with natural language definitions from IS literature:

- Containment: an experience of boundedness, entailing an interior, exterior and a boundary [159].

- Center_Periphery: the experience of objects or events as central, while others are peripheral or even outside [104]. The periphery depends on the center but not vice versa [187].
- Source_Path_Goal: a source or starting point, goal or endpoint, a series of contiguous locations connecting those two, and movement [159].
- Part_Whole: wholes consisting of parts and a configuration of parts [187].
- Support: Contact between two objects in the vertical dimension [211].
- Blockage: obstacles that block or resist our force; a force vector encountering a barrier and then taking any number of directions [159].
- On_Path_From: an object moves in a direction opposite to some entity [192]
- On_Path_Toward: an object moves in a direction toward some entity [192].
- Going_Through: an entity moves through a certain landmark [192, 319].

Examples of other frequently discussed IS are among others: Contact, Scale, Link, Balance, Object, Substance and Cycle. These IS are not yet in ImageSchemaNet because, compared to the above list, they are less documented in literature. This is due to various reasons. For example, Object is the abstraction of any sensorimotor experience of any bounded entity of the world, and this would result in considering all physical and, by metaphorical projection, all non-physical entities as activator of Object, which would not generate distinct, relevant knowledge. For other IS there is an ongoing debate on whether they should be considered IS in their own right, e.g. Contact is at times seen as a spatial primitive interacting with Support [214, 211] or its right as being an IS itself is debated [137]. Thus, the list above represents the most agreed upon selection of IS from literature at the moment of writing.

For a full and extended dissertation about IS history and contextualisation in embodied cognition theory, as well as for the description of ISL^M language in deep detail, refer to [135].

IS semi-automatic detection Related work in this direction has mostly focused on identifying image schemas in natural language by means of clustering verb-preposition pairs with noun vectors [116], also in a multilingual setting [115]. An extension of this traditional machine learning approach to include word embeddings has been proposed by Wachowiak [338]. One approach that relies on the Image Schema Repository [152], also used in the experimental setting of this work, is a fully automated method of classifying natural language expressions into image schema categories by fine-tuning a pre-trained neural language model [339]. While the results, especially of transferring the learned knowledge to other languages, are promising, there is still room for improvement. For instance, one short-coming of the previous approach is that it can only predict one image schema per natural language expression because of the nature of the ISCAT dataset, while multiple image schemas frequently co-occur in a natural language sequence.

Our work departs from previous research in formalising knowledge about IS in a modularized ontological structure: one module per each main broad pivotal contribution. Other than this, as mentioned in Chapter **??**, image schemas are here treated as frame structures, with semantic roles (spatial primitives) and lexical / factual triggers. The ontology developed adopting this approach is linked to Framester, and it operationalizes image schemas as a new layer on top of frame-based knowledge extracted from text. We include testing on a small evaluation dataset, using full-fledged Semantic Web techniques to design ImageSchemaNet. In contrast to clustering and neural approaches, the method for annotating natural language sequences with image schemas exposed in the following Section is fully explainable, since it keeps track of lexical units, their related frames, and the links between frames and formalized image schemas. Furthermore, we can identify more than one image schema per sequence when applicable.

2.2 ISAAC Ontology Network

ISAAC is the Image Schematic Abstraction and Cognition modular ontology. It is conceived as a broad spectrum hub of ontological modules related to studies on embodied cognition and image schemas. In this context, in particular, its aim is to formalize image schematic knowledge, considering different layers such as lexical triggers, factual knowledge, multi-lingual translations, conceptual matters, more complex constructs involving epistemic stance, attitude and polarity expression etc. The ISAAC ontology, furthermore, is integrated in the Framester hub and it introduces an embodied cognition knowledge layer on top of several semantic web resources integrated in the hub. Since a major flaw in current image schema theory is the lack of agreement about the lexical coverage of image schemas, we introduce an image-schematic layer linked to FrameNet [9], WordNet [225], VerbNet [286], etc., thereby creating a formal, lexicalized integration of cognitive semantics, enactive theories, and frame semantics. The main contributions of this approach are as follows:

- An image-schematic layer in the Framester hub called ImageSchemaNet that is easy to access by means of a SPARQL endpoint, linking image schemas to existing resources;

- A formal and re-usable representation of image schemas as Semantic Web technology in form of an ontological layer;
- An explicit representation of the interplay of existing (lexical) semantic and formal resources to interlink commonsense knowledge represented as image schemas to natural language and vice versa;
- An empirically evaluated method for automatically identifying image schemas in natural language sentences;
- An automatic, explainable frame-based detector of image schemas, that offer the possibility to reason over the graph structure including image schematic knowledge and its *locus* of activation.

In the following Sections we describe all of them in detail.

2.2.1 ISAAC module

The ISAAC ontology[1] consists of several modules, including: three main theoretical modules (i) *Johnson87* abbreviated to *J87*, presented in Sec. 2.2.2, modeled from [159]; (ii) a module on the work of Mandler and Pagán Cánovas [215] abbreviated to *MPC*, presented in Sec. 2.2.3; and (iii) Hedblom et al., abbreviated to *HED*, presented in Sec. 2.2.4, modeled from several works of Hedblom et al. [137, 136, 138]. Furthermore, the ISAAC modular ontology includes the ISCAT module, which is the ontological transposition of Image Schema Catalogue repository[2] (ISCAT) [151], presented in Sec. 2.2.5, and the ISFRAME module, which is the ontological transposition of the ISCAT repository including alignment to FrameNet [236] and MetaNet [97] frames, described in Sec. 2.2.6. Finally the ImageSchemaNet module, presented in Sec. 2.2.7 is the ontological module dedicated to the lexical and factual triggers, extracted using the QUOKKA workflow, to operationalize the three above mentioned theoretical modules; it allows the implementation of pattern graphs inferences, exposed in the experimental section 2.3.3, and it is used in Chapter 5 for joint experiments in order to extract knowledge from graphs automatically generated from natural language combined with values and emotions layers.

[1] The ISAAC ontology and all its modules are available here: https://github.com/StenDoipanni/ISAAC/tree/main/ISAAC_ontology_network

[2] Available, in a reworked version thanks to the work of Dagmar Gromann, here: https://raw.githubusercontent.com/dgromann/ImageSchemaRepository/main/IS_repository_final.csv

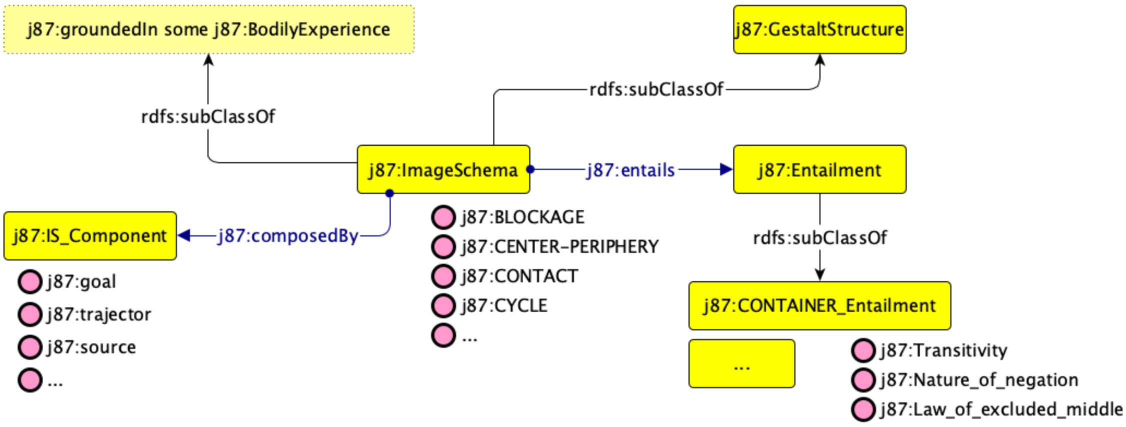


Figure 2.1: *J87* module basic structure.

2.2.2 Johnson 1987 (*J87*) module

J87 is the ontological module derived from "The Body in the Mind" [159], in particular from its Chapters 1-5, excluding chapters on imagination and a general theory of meaning. From here on we use `j87:` as prefix for all the entities modeled in the *J87* module.

Fig. 2.1 shows the basic structure of *J87* module. The ontological representation follows a top-down approach and the main notable classes are the following:

- `j87:GestaltStructure`: 'Gestalts' are described as "not unanalyzable givens or atomistic structures. They can be 'analyzed', since they have parts and dimensions, but any such attempted reduction will destroy the unity (the meaningful organization) that made the structure significant in the first place" [159].

- `j87:ImageSchema`: this class is annotated, via Exuviae `ex:bibRef` annotation property, with all the different definitions provided in the source publication. Image schemas are explicitly defined as "Image Schematic Gestalt Structures", therefore this class is modeled as subclass of `j87:GestaltStructure`. Furthermore, according to Johnson [159], the `j87:ImageSchema` class is restricted as being the subclass of something which is `j87:groundedIn` some `j87:BodilyExperience`.

- `j87:Entailment`: this class represents a cluster of possible, probable, necessary or prototypical implications that an image schema might have. Johnson does not formalize or operationalize the following assumptions, therefore they have to be intended as top-down theoretical assertions without any ambition of logical formalism. For instance in "The Body in the Mind" [159] CONTAINER is described as being the schema for three main entailments, which are modeled in *J87* as `j87:CONTAINER_Entailment` instances: `j87:Law_of_excluded_middle`, `j87:Transitivity`, and `j87:Nature_of_negation`. It is clear that even if, for the sake of coherence with the original source, they are represented at the same ontological level, these individuals are very different entities. In fact, `j87:Law_of_excluded_middle` states that everything is either P (in the container) or not-P (outside the container). `j87:Transitivity` is said to be entailed by CONTAINER for the prototyping of the syllogism:

$$containedIn(A,B) \quad \wedge \quad containedIn(B,C) \quad \Longrightarrow \quad containedIn(A,C)$$

based on the assumption that the containment relation is usually transitive, thus referring to transitivity of containment.
Finally, `j87:Nature_of_negation` is instead listed because Johnson states that our ability to negate derives directly from our conceptualisation of categories as containers, so negating something is stating that it is not in some category and therefore, it is not *containedIn* some (metaphorical) container.

- `j87:GestaltCriterion`: this class takes as instances those that are said to be the necessary criteria for a gestalt structure to be "emergent and salient" in our experience, and these criteria are `j87:Pervasive`, `j87:Simply-structured`, `j87:Well-structured` and `j87:Well-understood`. However, these criteria to recognise an image schema are neither formalised nor defined consistently.

Image schemas are modeled as subclass of gestalt structures, having "parts and dimensions" [159]. The relation of subsumption is utilised when an image schema is portrayed to be more specific than another, e.g. BALANCE subsumes AXIS_BALANCE, EQUILIBRIUM, POINT_BALANCE and TWIN_PAN_BALANCE. Considering the relational structure of this module a form of compositionality among image schemas is expressed by Johnson about PATH, for which the property `dul:hasComponent` is used to state PATH `dul:hasComponent` SOURCE, force_vector, GOAL and vector_tracing_a_path.[3] The *J87* module, as well as further documentation, is available on the ISAAC GitHub repository[4].

***J87* Competency Questions** *J87* module allows to answer some CQs according to *J87* theory, such as:

1. What are the entailments for some image schema?

2. Are all image schemas also gestalt entities?

[3] No graphical notation is used for some of these image-schematic structures since no precise type or exact nature is provided by Johnson, so using the graphical notation would mean committing to an idea without theoretical nor empirical grounding.

[4] The *J87* module is available here: https://github.com/StenDoipanni/ISAAC/blob/main/ISAAC_ontology_network/j87.owl

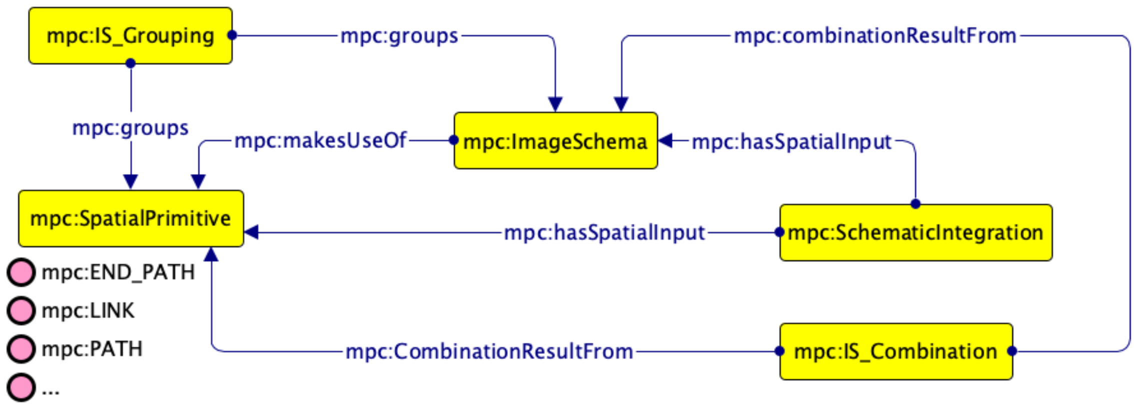


Figure 2.2: *MPC* module basic structure.

3. Does all the image schemas have some IS component?

The SPARQL queries for these competency questions are available on the ISAAC GitHub repository[5].

2.2.3 Mandler and Pagán Cánovas (*MPC*) module

Mandler and Pagán Cánovas [215] introduce crucial terminological and methodological distinctions taken up by many other works, especially that of spatial primitive and image schema. From here on we use `mpc:` as prefix for all the entities modeled in the *MPC* module. *MPC* introduces the `mpc:SpatialPrimitive` class that represents image-schematic conceptual building blocks.

MPC specifies Johnson's vague "components" as spatial primitives and shifts the angle from a linguistic and philosophical to a psychological developmental perspective. This is represented in *MPC* module by the entity `mpc:DevelopmentalPerspective`, which is an instance of the `mpc:IS_Approach` class, modeling the "Image Schema Approach" of this module.

***MPC* Classes** The *MPC* module includes the following main classes:

- `mpc:SpatialPrimitive`: entity modeling those elements described as the "first conceptual building blocks formed in infancy" [215]. They are described as roles for image schemas;

- `mpc:ImageSchema`: described as "simple spatial stories build from spatial primitives" [215];

[5] Queries and additional material are available here: `https://github.com/StenDoipanni/ISAAC/tree/main/ISAAC_ontology_network`

- mpc:SchematicIntegration: blending of spatial primitives or image schemas with non-spatial elements, e.g. values, emotions, etc.

- mpc:IS_ComplexityCriterion: first preverbal conceptual understanding of infants and its development provide information on the most fundamental image schemas, i.e., the time of development of conceptual understanding specifies the complexity of the associated image schema;

- mpc:IS_Combination: has instance mpc:ANIMATE_THING, which combines mpc:ANIMATE_MOVE and mpc:THING, however, which parts/qualities/roles are combined or whether it represents a coactivation of distinct image schemas remains unclear;

- mpc:IS_grouping: clustering image schemas in groups.

***MPC* Object Properties** To declare relations among its classes the *MPC* module introduces important object properties. For instance, mpc:hasSpatialInput with domain mpc:SchematicIntegration and range mpc:SpatialPrimitive or mpc:ImageSchema, is defined as a "process similar to what is called simplex network in Conceptual Integration Theory" [215]. It refers to the role that some unstructured element takes from the organizing frame structure imported from the input space. The inverse relation to this is mpc:isTopologyProviderFor that, even without being a direct commitment, implies a similarity between image schema components and roles of a frame. The main commitment of this module is the mpc:makesUseOf object property, that has range mpc:SpatialPrimitive and makes explicit the 'role-frame'/'spatial primitive-image schema' parallel suggested by Johnson's definition of IS as gestalt structures. The *MPC* module is available on the ISAAC GitHub[6].

***MPC* Competency Questions** *MPC* module allows to answer some CQs according to MPC theory, such as:

1. What image schema makes use of what spatial primitive?

2. What image schemas are grouped in what IS grouping?

3. How many are all the spatial primitives?

[6]The *MPC* module is available here: https://github.com/StenDoipanni/ISAAC/blob/main/ISAAC_ontology_network/mpc.owl

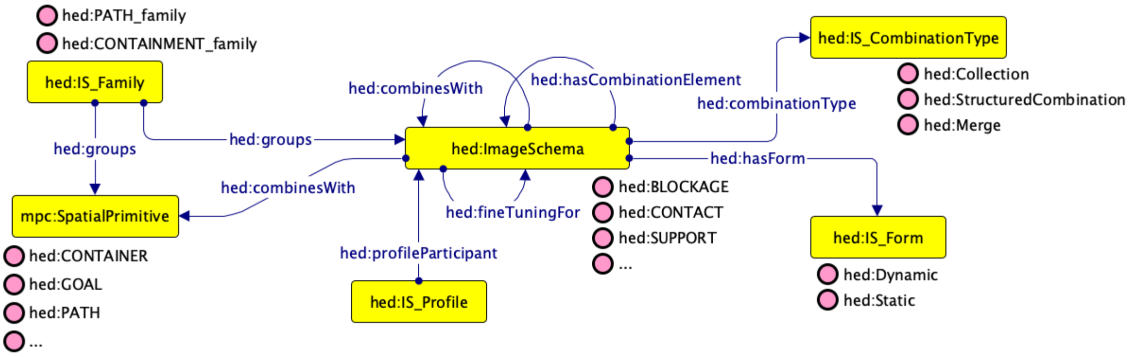


Figure 2.3: *HED* module basic structure.

The SPARQL queries for these competency questions are available on the ISAAC GitHub repository[7].

2.2.4 Hedblom (*HED*) module

The *HED* module inherits the theoretical approach from the *MPC* one, and focuses on representing all the image schemas and spatial primitives specified in the listed publications. It provides a sound system for IS compositionality. From here on we use `hed:` as prefix for all the entities modeled in the *HED* module. Object properties and SWRL rules are used to express axioms originally represented in the DOL language [138].

***HED* Classes** One of the most central theoretical contributions in the *HED* module is `hed:IS_Family`, which attributes a specific name to the clustering of image schemas, in contrast to the generic `mpc:IS_grouping`. It explicitly declares an additional dimension: complexity. In contrast to the complexity criterion in *MPC*, it focuses on the complexity of compositionality, rather than the complexity of individual image schemas. Two of the main classes are `hed:SpatialPrimitive` and `hed:SpatialSchema`, equivalent to the *MPC* notion of image schemas and spatial primitives.

Here the main classes used to formalize the introduction of new formal constraints:

- `hed:IS_Complexity`: Image Schema Complexity increases proportionally to the addition of spatial primitives. The assertion of some IS as 'more complex' than some other, is realized in OWL 2 via the object property `hed:fineTuningFor`. For more information see Hedblom [138].
- `hed:IS_Family`: Image schemas consist of different "parts"; These parts can either be removed or added while still capturing the same basic image schema, generating

[7] Queries and additional material are available here: `https://github.com/StenDoipanni/ISAAC/tree/main/ISAAC_ontology_network`

what can be described as an image schema family [138].

- `hed:IS_Form`: IS are here presented in two possible forms: static or dynamic. The "static" form denotes the notion of the possibility of things being in some way, i.e., the spatial co-location of entities and its configuration. The "dynamic" form equally denotes the co-location of entities with the addition of some form of movement of one or both entities [136].

- `hed:IS_profile`: The IS profiles are defined by Oakley as "groupings of image schemas that capture the spatiotemporal relationships related to particular events" [237], namely e.g. the set of IS activated by some concept or event, or even sentence, description, situation etc.

- `hed:IS_CombinationType`: this class is used to express the possible combination types, which is subclass of `hed:IS_Transformation` and takes as instances three types of combinations: `hed:Collection`, `hed:StructuredCombination`, and `hed:Merge` as described in Sec. 2.1, and originally in Hedblom et al. [139].

***HED* Object Properties** The relations between IS, SP, and other entities are formalized via object properties. Here we list the most relevant:

- `hed:combinationType`: some IS has some combination type.

- `hed:combinesWith`: some IS can combine with some other IS or with some SP to compose more complex structures.

- `hed:fineTuningFor`: This property expresses the increasing complexity of IS through the use of one or more SP. Some more basic IS use just one or few SP, while more complex ones are "fine tuning" the more basic ones, since they are more specialised and articulated in their compositional structure and meaning (e.g. linear path movements vs circular path ones) [138].

- `hed:groupedInFamily`: some IS or SP is grouped in some IS family. Its inverse is `hed:groups`.

- `hed:hasCombinationElement`: some IS has combination element some other IS. More specificity, according to the combination type, can be expressed with its three subproperties: `hed:hasCollectionElement`, `hed:hasSequenceElement` and `hed:hasMergingElement`.

- `hed:hasForm`: some IS has form static or dynamic.

- hed:IS_familyIntersectionWith: two IS families can have an intersection when some IS is grouped in (at least) two families e.g. hed:GOING_IN is the intersection result of hed:SOURCE_PATH_GOAL and hed:CONTAINMENT.

- hed:IS_profile: this property links some dul:Concept or dul:Event to some IS profile, in turn the IS profile has some hed:profileParticipant some IS.

- hed:makesUseOf: some IS makes use of some SP or some other IS.

Furthermore, the *HED* module includes several SWRL rules to operate (i) property value assignment and (ii) named individual inference rules, not possible with OWL expressivity.[8] All the following SWRL rules are expressed in SWRL syntax.

While extended description of the transposition from DOL to OWL 2 is explained in the ISL2OWL GitHub[9], here some examples are provided. The following SWRL rules also exemplify a good portion of the competency questions that this module allows. The SPARQL queries to test the following SWRL rules are available on the ISAAC GitHub repository[10].

The object property hed:fineTuningFor derives from Hedblom [138], and it is used to relate a more complex IS to a less complex one e.g. SOURCE_PATH_VIA_GOAL is a more complex version (fine tuning) for Source_Path_Goal. A more complex version of some IS makes use of (at least) the same SP as its less complex version. This axiom is formalized in *HED* module as a SWRL rule:

$$\texttt{hed:makesUseOf(?x,?y)} \;\wedge\; \texttt{hed:fineTuningFor(?z,?x)} \rightarrow \texttt{hed:makesUseOf(?z,?y)} \tag{2.1}$$

Asserting that: if some entity ?x, (an image schema) hed:makesUseOf some ?y, (a spatial primitive), and some entity ?z is declared as being more complex than ?x via the property hed:fineTuningFor, then ?z hed:makesUseOf the same spatial primitives of ?x. Another distinction among IS introduced in this module, not always clear on the linguistic level, is the hed:IS_Form. The hed:IS_Form differentiate hed:Static IS from hed:Dynamic IS. SWRL rules 2.2 and 2.3 are about hed:Static and hed:Dynamic IS form:

[8] All the SWRL rules inferences are developed with the SWRLTab 2.0.11 Protégé-OWL development environment and tested with both HermiT 1.4.3.456 and Pellet reasoners in Protégé, version 5.5.0.

[9] The ISL2OWL repository is available here: https://github.com/StenDoipanni/ISAAC/tree/main/ISL2OWL

[10] Queries and additional material are available here: https://github.com/StenDoipanni/ISAAC/tree/main/ISAAC_ontology_network

$$\texttt{(hed:hasForm(?x,hed:Static)} \wedge \texttt{hed:isStaticFormFor(?x,?y))} \rightarrow \texttt{hasForm(?y,hed:Dynamic)} \tag{2.2}$$

$$\texttt{(hed:hasForm(?x,hed:Dynamic)} \wedge \texttt{hed:isDynamicFormFor(?x,?y))} \rightarrow \texttt{hasForm(?y,hed:Static)} \tag{2.3}$$

In detail Axiom 2.2 states that if some `?x` `hed:hasForm` `hed:Static` and `hed:isStaticFormFor` some `?y`, than `?y` `hed:hasForm` `hed:Dynamic`. Axiom 2.3 is the opposite of Axiom 2.2.

This module also provides a class for the grouping of image schemas activated when conceptualizing a complex event, action, etc., that is a `hed:IS_profile`. One example taken from [139] is the "turducken"[11]. The `hed:Turducken_profile` is an instance of `hed:IS_profile`, taking as `hed:profileParticipant` the image schema GOING_IN, SCALE, ITERATION and CONTAINMENT. A further novel element in the *HED* module is the class `hed:IS_CombinationType`, subclass of `hed:IS_Transformation` and taking as instances three types of image schema combinations [139], namely, `hed:Collection`, `hed:Merge` and `hed:StructuredCombination`. Axioms formalizing these three type of combinations and allowing further useful inferences are:

$$\texttt{hed:hasMergingElement(?x,?y)} \rightarrow \texttt{hed:combinationType(?x,hed:Merge)} \tag{2.4}$$

$$\texttt{hed:hasCollectionElement(?x,?y)} \rightarrow \texttt{hed:combinationType(?x,hed:Collection)} \tag{2.5}$$

$$\texttt{hed:hasSequenceElement(?x,?y)} \rightarrow \texttt{hed:combinationType(?x,hed:Sequence)} \tag{2.6}$$

$$\texttt{hed:hasCombinationElement(?x,?y)} \wedge \texttt{hed:hasCombinationElement(?x,?z)} \wedge \texttt{owl:differentFrom(?y,?z)} \rightarrow \texttt{hed:combinesWith(?y,?z)} \tag{2.7}$$

[11] A dish with a chicken stuffed inside a duck that in turn is stuffed inside a turkey.

$$\texttt{hed:combinesWith(?x,?z)} \wedge \texttt{hed:groupedInFamily(?x,?y)} \wedge$$
$$\texttt{hed:groupedInFamily(?z,?k)} \wedge \texttt{hed:hasMergingElement(?h,?x)} \wedge$$
$$\texttt{hed:combinationType(?h,hed:Merge)} \rightarrow \texttt{hed:IS\_familyIntersectionWith(?y,?k)} \quad (2.8)$$

Axiom 2.4 states that if some `?x` `hed:hasMergingElement` some `?y`, then the combination type of `?x` is `hed:Merge`. Axioms 2.5 and 2.6 state the same about the other two types of combination: `hed:Collecion` and `hed:Sequence`. Axiom 2.7 allows the inference that, given an entity `?x` and a number N of elements `?y`, if `?x` `hed:hasCombiningElement` more than one element, then each of them `hed:combinesWith` all the others. Finally, Axiom 2.8 formalizes IS families intersection: if `?x` `hed:combinesWith` `?z` and `?x` is `hed:groupedInFamily` `?y`, while `?z` is `groupedInFamily` `?k`, and an entity `?h`, being the result of a merging process, `hed:hasMergingElement` `?x`, then we can say that there is an occurrence of a `hed:IS_familyIntersectionWith` `?y` (`?x`'s family) and `?k` (`?y`'s family).

The Hedblom module, building on the conceptualization of moving objects [215], addresses spatial primitives' compositionality, whose addition or subtraction determines the structure of a specific image schema (e.g. MOVEMENT_ALONG_A_PATH + START_PATH = SOURCE_PATH).

The *HED* module is available on the ISAAC GitHub[12].

2.2.5 ISCAT

The ISCAT module is the ontological transposition of the ISCAT repository. The main class is `iscat:ConceptualMetaphor`, while each conceptual metaphor from the original dataset is represented as an individual, while each lexical example is the object of the data property `iscat:hasExample`. The resource presents examples from English and German, specified in round brackets before the example string object of the triple, while the original resource from which the example is taken is in squared brackets, after the example string. To provide an example, a triple expressed in turtle syntax about the cognitive metaphor *Morality is a straight path* is:

[12] The *MPC* module is available here: `https://github.com/StenDoipanni/ISAAC/blob/main/ISAAC_ontology_network/hed.owl`

$$\begin{aligned} \texttt{iscat:MORALITY\_IS\_A\_STRAIGHT\_PATH} \quad \texttt{iscat:hasExample} \\ \text{“}(English)\ He\ is\ a\ deviant.\ [MetaNet]'', \\ \text{“}(English)\ He\ deviated\ from\ the\ straight\ and\ narrow.\ [MetaNet]'', \\ \text{“}(English)\ She\ strayed.\ [MetaNet]''. \end{aligned} \tag{2.9}$$

Some of the main resources from which examples are taken are the MetaNet repository [63], Grady [109], Jäkel [155] Lakoff & Johnson [190], [191], Baldauf [11], Kimmel [174], and Santibanez [280].

2.2.6 ISFRAME

The ISFRAME module imports the ISCAT one, and adds the source and target domain of the conceptual metaphor, disambiguated on frames from the FrameNet and MetaNet resources. The object properties introduced in this module are:

- `isframe:hasSourceFrame`: it is used to declare the frame evoked by the source domain used in the metaphor;
- `isframe:hasTargetFrame`:it is used to declare the frame evoked by the target domain used in the metaphor;
- `isframe:gStructBy`: it is the property used to declare which is the the image schema that is “gestaltically structuring” the metaphor.

To provide an example, considering the same metaphor from the Section above, the information introduced in this module, expressed in Turtle syntax, is as follows:

$$\begin{aligned} \texttt{iscat:MORALITY\_IS\_A\_STRAIGHT\_PATH} \\ \texttt{isframe:hasSourceFrame} \quad \texttt{metanet:Straight\_path} \quad ; \\ \texttt{isframe:hasTargetFrame} \quad \texttt{metanet:Morality} \quad ; \\ \texttt{isframe:gStructBy} \quad \texttt{hed:SOURCE\_PATH\_GOAL} \quad . \end{aligned} \tag{2.10}$$

As shown in axiom 2.10 the conceptual metaphor is imported from the ISCAT module, the properties are introduced from anew in the ISFRAME module, the original resources are reused from the MetaNet repository, while the image schema is reused from the HED module.

2.2.7 ImageSchemaNet

ImageSchemaNet relies on ISAAC, the Image Schema Abstraction And Cognition ontology. ISAAC models both formal and semi- or unstructured state-of-the-art IS theories, and proposes an integrated theory combining Johnson's definition [159] of image schemas as gestalt structures, Mandler and Pagán Cánovas spatial primitives conception [214] as "first conceptual building blocks", and Hedblom's IS compositionality [22]. ISAAC uses Framester (and derivatively Fillmore's Frame Semantics) to deliver a reified representation of situations evoked in natural language as occurrences of frames and their foundational IS.

ImageSchemaNet reuses the `ex:bibRef` property from Exuviae [50], described in Sec. 1.4, which is meant to keep precise reference of the bibliographical and theoretical provenance of each entity and property with the original definition and formal dependencies. In particular, ImageSchemaNet focuses on the `is:ImageSchema`, `is:SpatialPrimitive` and `is:IS_Profile` classes from the ISAAC ontology, and introduces the `:activates` property in order to declare assertions about the activation (i.e., a bodily-schematic evocation) of some image schema or spatial primitive from any entity in the Framester resource.

The ImageSchemaNet ontology is available and can be queried from the Framester endpoint[13]. A detailed documentation about the structure, querying, and evaluation is provided in the following sections, as well as on the ImageSchemaNet GitHub repository[14]. Albeit importing ISAAC ontology, ImageSchemaNet specifically focuses on providing lexical coverage to the Image Schema Theory, via `isn:activation` assertions, which currently cover the following image schemas: Containment, Center_Periphery, Source_Path_Goal, Part_Whole, Blockage, and Support.

2.2.8 ImageSchemaNet classes

`is:ImageSchema` The `is:ImageSchema` class represents the general concept of Image Schema, it is defined using the `ex:bibRef` property, quoting literature definitions, and it takes as instances image schemas whose activation is covered in the ImageSchemaNet ontology. Each IS is axiomatized as a gestalt structure, composed by at least 2 spatial primitives, and it is modeled as a kind of conceptual frame.

`is:SpatialPrimitive` The `is:SpatialPrimitive` class takes as instances the "first conceptual building blocks formed in infancy" as in [215], and represents them as semantic roles. The labels used respectively for IS and SP refers to well established and documented

[13] `http://etna.istc.cnr.it/framester2/sparql`
[14] `https://github.com/StenDoipanni/ISAAC/tree/main/ImageSchemaNet`

names used in literature, as for the Support IS, quoting their definition and provenance. When specific “official” names were not already given to entities, which existence was nonetheless implicitly or explicitly stated, we used labels extracted from empirical use case. For instance, in the aforementioned Support case, where literature is often mentioning examples involving its spatial primitives, no official name was available, and for this reason the Supporter and Supported SP were introduced.

`is:IS_Profile` The `is:IS_Profile` class is used as in [139] and [237] to describe the collection of IS which are activated by some entity, sentence, situation or event. One of the relevant future developments stemming from our work is the automatic extraction of the image schema profile and the investigation of the conceptual nature of relations among IS in such a collection. The prominence of one particular IS in a set generated from, for example, a text string, refers to a form of frame compositionality as in [89], which could be determined by syntax as well as discourse structure, depending on term, sentence and text compositionality. The `is:IS_Profile` class is particularly relevant here since it’s the class used for our evaluation system as described in Section 2.3.3.

2.2.9 ImageSchemaNet properties

All the activation declarations in ImageSchemaNet are realized via the `is:activates` object property or its subproperties, which specify details about the way, layer, resource and type of activation. The meaning of `is:activates` refers to some element that *activates* the cognitive substratum that is associated with an image schema. For instance, the verb *to contain*, the noun *container*, the frame `Containment`, and the frame element `Container` all activate the image schema Containment.

For this reasons, the following sub-properties were introduced in the graph:

- `is:activates` : declares the activation from a Framester or Framenet frame to an IS. It is the super-property to all the following properties;

- `is:closeMatchActivation` : used for the activation of some IS from entities which have a `skos:closeMatch` (close alignment declarations from Framester) to a FrameNet frame that activates an IS;

- `is:coreSPActivation`, `is:peripheralSPActivation`, `is:extraThematicSPActivation` : used for the activation of spatial primitives from FrameNet frame elements, which are distinguished into core (necessary), peripheral (optional), and extra-thematic (not frame-specific);

- `is:lexicalSenseActivation` : used for lexical entities directly evoking spatial primitives or image schemas. This property represents activation based on: 1) very accurate manually verified alignments; 2) alignments inferred from logical rules. For example, the IS activation from WordNet synsets and Framester frames follows this heuristic: if a synset *s* evokes (in Framester semantics is subsumed by) a frame *f*, and the frame *f* activates an IS *i*, then *s* activates *i*;

- `is:semTypeActivation` : used for semantic types used e.g. in FrameNet or VerbNet as selectional restrictions, which activate image schemas or spatial primitives;

- `is:semanticRoleActivation` : used for VerbNet arguments, FrameNet frame elements and PropBank roles activating spatial primitives;

- `is:gestaltActivation` : activation of an image schema through its spatial primitives. Due to their "gestalt nature"[164], the activation of a spatial primitive implicates the activation of the whole schema.

Section 2.2.10 describes the most productive steps of image-schematic triggers knowledge graphs population applying the QUOKKA workflow. It provides example of some fallacies, especially in its automatic steps, including interesting case for which the IS activation declaration would be debatable. Furthermore, albeit the existence of the above mentioned specific subproperties, kept in order to understand the semantics of the activation of some entity, the online version of the ImageSchemaNet as well as the graphs produced with the automatic detector, directly use the `is:activates` not dependently on the type of the entity object of the triple. This was done to facilitate the exploration of the resource and to prioritize applicability at the expense of unnecessary granularity at this stage of the work. Section 2.3 describes the Image Schema detector tested on the ISCAT repository, while experiments involving image schemas, values and emotions are described in Chapter 5.

Some useful queries to explore the ImageSchemaNet resource, using the aforementioned properties can be found on the ISAAC repository[15].

2.2.10 Grounding ImageSchemaNet

This Section describes the QUOKKA workflow, introduced in Sec. 1.5, applied to the domain of image schemas. There is no repository that aligns entities from different semantic layers (lexical units, semantic roles, framal structures, factual entities, etc.), to

[15]Queries to explore the resource are available here: `https://github.com/StenDoipanni/ISAAC/blob/main/ImageSchemaNet/imageschemanet_etr_queries.txt`

image schemas and spatial primitives. Moreover, while a few references could be found in FrameNet, no lexical grounding has been provided for image schemas.

Since ImageSchemaNet is an extension of Framester, and image schemas are represented as a special kind of frames activated by other Framester elements, that grounding is straightforwardly performed according to the heuristic abstraction method presented in the following. We firstly provide a simple example of how ImageSchemaNet can be used after being grounded, in order to make the process more intuitive to the reader.

Consider the sentence: *The Obama administration had entered into an agreement with Iran*, we can: (a) tokenize the sentence into its main elements (*Obama administration*, *enter into agreement with*, *Iran*), (b) collect their senses and (c) disambiguate the contextually valid ones (e.g. Obama_Presidency from DBpedia entity linking, Enter_51010000 from VerbNet disambiguation, Joint Comprehensive Plan of Action from DBpedia, Iran State from DBpedia), (d) retrieve the frames evoked by the senses (Organization, Path_shape, Be_in_agreement_on_action, Political_locales, all from FrameNet), and finally (e) retrieve the image schema activated by an entity, a sense, or a frame (Organization = nil, Path_shape = Source_Path_Goal, Be_in_agreement_on_action = nil, Political_locales = nil)[16].

In practice, the heuristic abstraction method reveals that the main image schema activated by the sentence is Source_Path_Goal. The inferential structure of Source_Path_Goal can further lead us to infer the roles played by an organization, an observed situation, and a political locale.

The exemplified heuristic abstraction can be performed with automated tools, which are evaluated in Section 2.3.4. Here the focus is on presenting the hybrid grounding procedure used to populate ImageSchemaNet on top of Framester. We have used the queries listed in ImageSchemaNet repository, following the workflow described in Sec. 1.5, together with manual revision.

The queries can be reproduced on the Framester endpoint by substituting (manually or programmatically) the `insert_variable` element with the corresponding entity, as specified in the query description, and by providing the correct prefix. Figure 2.4 shows the most productive steps as well as relations among IS/SP activators in different resources in Framester. Classes of entities activating IS/SP, represented as yellow rectangular boxes, are retrieved via SPARQL queries, represented as blue oval shapes, and described in the next sections.

Each query returns a different number of results depending on the entity introduced as `input_variable`. Figure 2.4, shows how some entities, being retrieved by some query,

[16] The “nil” values could be further populated by looking for possible activated spatial primitives as well as other knowledge layers e.g. emotions and values

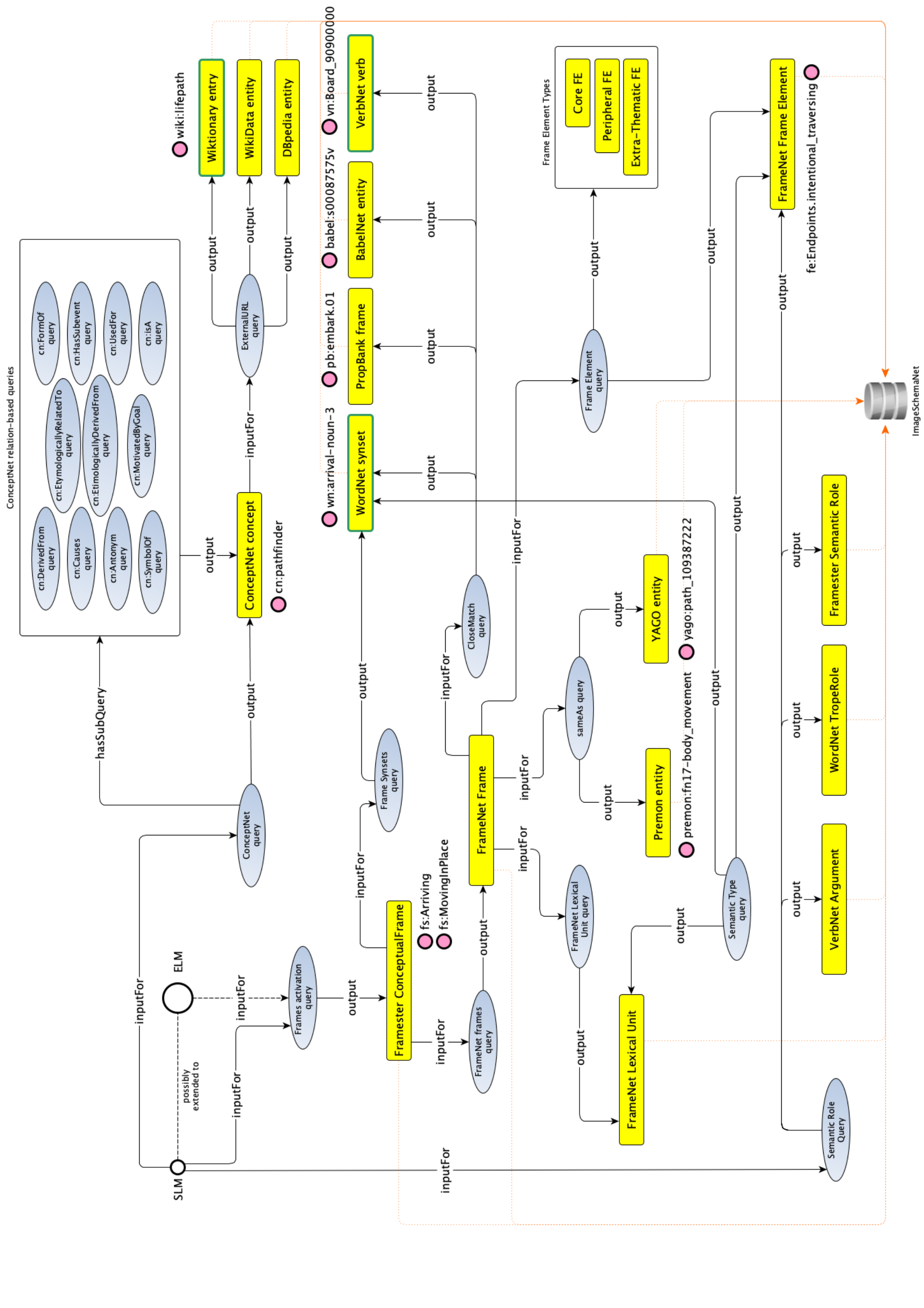


Figure 2.4: QUOKKA frame building workflow applied to Image Schemas domain, here showing some IS activators for SOURCE_PATH_GOAL.

are used as input for other queries. Rectangular boxes with no incoming `output` and only `inputFor` arrow represent those steps that need a human in the loop e.g. the SLM starting step, or the "semantic type" query, in order to produce meaningful results, require some domain expert which analyzes all results and filters them manually. All the other steps can be automatized, although, due to the great amount of knowledge in Framester resource, a manual check could result in higher quality data. Table 2.1 shows some data about ImageSchemaNet KG population from different resources. Details are discussed in the following paragraphs.

Frame-driven activation We have started looking for the frames activating an IS. The first search uses a non-disambiguated lexical unit e.g. *contain* for the Containment IS, as Starting Lexical Material (SLM) to retrieve all the senses and frames evoked by a lexical unit in isolation. For example, for *contain*, the searching process can collect all its senses, and their evoked frames. Based on sense inheritance hierarchies (as available in OWL versions of WordNet and other lexical resources), the search is extended to more specific or more generic senses of e.g. *contain*, so extending the set of evoked frames, and potentially activated IS. This kind of query is exemplified on Framester and can be found in ImageSchemaNet repository as "Frames Activation Query", in Figure 2.4 as "Frame activation query" node, and in the OWL file as annotation of the `:activates` object property using the `:operationalizedVia` annotation property.
However, the amount of senses and related frames can be large, and we need *contextual disambiguation* in order to make it more precise. As shown from Tab. 2.1 the highest amount of frames conceptually grounded in an image schema are by far related to dynamics of *movement*, making Source_Path_Goal the biggest among all the IS frame graphs. Note that, being On_Path_From, On_Path_Toward and Going_Through specifications of the Source_Path_Goal IS, they are treated as subframes, namely all the frames that trigger a specific type of movement e.g. moving towards / far from some entity, also activates the general idea of movement along a path. After performing the query, the selection of frames activating an IS is done manually, and after the iteration of the query for all synonyms and hyponyms, the first phase of frames activation search is closed, and we move to the frame element activation search.

Frame element-driven activation Frame element activation concerns the activation of an image schema or spatial primitive, and can be performed similarly as with frames. This kind of query is exemplified by focusing on retrieving FrameNet frame elements of type "Core", "Extra-Thematic" and "Peripheral". After performing the query, a further selection of frame elements is done manually, by using as pivotal the set of frames selected

Table 2.1: Image Schemas and amount of triggers from each semantic web resource involved in the emotion frame building process.

Image Schema	Frame	Frame Element	WordNet	VerbNet	PropBank	ConceptNet	Wiktionary	WikiData	DBpedia	BabelNet	Umbel	YAGO	Premon
CONTAINMENT	5	95	2292	203	81	39	41	2	2	1611	2	1314	5
SUPPORT	2	27	268	38	10	36	45	4	4	148	9	64	2
CENTER_PERIPHERY	4	32	685	80	20	57	55	1	1	466	4	339	5
BLOCKAGE	11	96	671	148	62	93	100	1	1	449	2	233	7
CONTACT	5	58	979	241	53	67	68	3	3	553	1	287	5
LINK	2	47	702	137	44	67	68	2	2	410	1	266	3
PART_WHOLE	6	33	272	45	47	74	82	2	2	181	2	88	3
SOURCE_PATH_GOAL	36	419	3716	882	955	131	130	1	1	2701	6	1075	26
ON_PATH_TOWARD	3	50	547	80	92	0	0	0	0	397	0	195	3
ON_PATH_FROM	5	91	868	237	142	34	29	2	2	594	0	173	5
GOING_THROUGH	2	57	200	52	30	0	0	0	0	126	0	34	2

in the step before, possibly enriching the set with further frames, not retrieved by the "Frames Activation query". Note that there is no direct correlation between the amount of frames and the amount of frame elements activated by each IS. This is due to two main reasons: the first reason is that some frames are highly structured, involving more than 10 semantic roles per each, other FrameNet frames are classified as "Non-Lexicalized", namely, no role nor lexical unit is indicated as evoking them, for their being considered "purely conceptual". In the FrameNet semantics this makes sense, but we tried to fill this gaps via using several semantic web resources, which cover different knowledge and conceptualisation layers. The second reason is that some roles can be included as IS triggers despite their frame is not directly related to some IS. Let's consider for example the `fn:AttemptSuasion` frame: it is not directly grounded in any IS, but one of its non-core roles, the `fn:Purpose.AttemptingSuasion`, for being conceptualised as a Goal it could be declared as activator of the Goal spatial primitive.

Lexical unit-driven activation Activation from lexical material is a substantial part of the heuristic abstraction, and it is generated by automatically querying Framester knowledge base, asking for all the elements (typically WordNet synsets or VerbNet verb senses) that evoke a frame. The query is performed for all the frames retrieved and selected as activators by the Frame Activation query. The heuristic rule here is: if an entity evokes a frame, which activates an IS, than that entity should have some form of activation for the IS too. The amount of elements retrieved may be considerable (for some IS, thousands of WordNet synsets), depending on the lexicalisation of the original frame. As a consequence, the synsets in the knowledge base are very useful for the coverage they provide in the populated ImageSchemaNet knowledge graph.

Of course, this coverage may contain some noise, since synsets are retrieved by making an inference from previous existing alignments in Framester, which may have different levels of confidence on their turn (there is a manually curated kernel of alignments, while other sets have used various heuristic rules, we refer back to section 1.5 and to QUOKKA repository for queries including the property `fs:hasReliabilityScore` to control the quality of alignments). To demonstrate the conceptual problems that have to be faced: both the terms "vase" and "absolutism" activate the Container image schema. This happens due to the fact that theoretical concepts of philosophical doctrines or behavioural attitudes (e.g. "absolutism") are aligned with a `skos:closeMatch` to the FrameNet frame `Containing`. In practice, "absolutism as a container" could be considered valid only when conceptual metaphors, e.g. IDEAS ARE OBJECTS, THINKING IS OBJECT MANIPULATION or CATEGORIES ARE BOUNDED REGIONS, are taken into account. A part of Section **??** is dedicated to further other critical examples. This query could be found on the QUOKKA GitHub repository as “Lexical Elements Activation Query”.

Semantic role-driven activation Activation assertions to FrameNet frame elements are extended through the multiple sources of semantic roles present in Framester (VerbNet arguments, PropBank roles, WordNet tropes, etc.). Semantic roles in Framester are organized as a complex taxonomy with a small top level that helps integrating them, and getting to the activated IS. The activation of spatial primitives, modelled as semantic roles, is materialised via the `:semantiRoleActivation` property. Roles are retrieved with two queries, starting from top nodes of different graphs, in order to declare the activation of both general and specific roles. The queries are available on the QUOKKA GitHub repository under the label of “General Semantic Roles Activation” and “Specific Semantic Roles Activation”.

Semantic type-driven activation A final important aspect is constituted by the FrameNet semantic type of entities. For example, FrameNet semantic types `fnst:Lateral` and `fnst:Leftish` activate Center_Periphery, while the frame element `Goal` in frames like `Attaching`, `Body_movement` or `Bringing` has the FrameNet semantic type `fnst:Goal`, which activates the Goal spatial primitive. Note that the semantic type activation refers back to Kuhn’s work [184] of “affordances”. The semantic type `fnst:Building`, including all types of building, having the teleological nature of containing entities, activates the Containment IS.

Further examples are provided in Section 2.3.4. The queries used for semantic type activation assertions include an initial query listing all existing semantic types, followed by a manual exploration of their differences and coverage, resulting in a selection of semantic types activating some IS or SP. Then, a second query is performed, looking for entities filtered by the aforementioned iteration of non-disambiguated lexical units from synsets and their hyponyms, also extracting their semantic type, ending in a final coherence checking between the entities retrieved, their semantic type, and their semantic type activation of an IS or SP. The query is shown in Fig. 2.4 as “Semantic Type Activation Query”.

2.3 ImageSchemaNet Evaluation

Devising an evaluation method for ImageSchemaNet is not an easy task, since there is no available formal resource featuring IS activation, and no tool able to automatically detect and extract IS from text. Due to this cold start problem, no baseline is proposed. Starting from a corpus of manually annotated sentences, we have performed an evaluation of ImageSchemaNet as an extension to existing automated methods: the end-to-end OpenSesame frame parser, and the hybrid FRED frame-based machine reader, as described in Sec. 1.6.1. In practice, we have taken an existing corpus of sentences, manually annotated with IS and SP, and we have measured the accuracy of automatically inferring IS based on mapping frames – detected by OpenSesame and FRED – to IS/SP.

2.3.1 Evaluation setting

The evaluation setting uses an excerpt of the ISCAT dataset[17] [150], and state-of-the-art tools for frame detection from natural language. The ISCAT excerpt has been taken from a cleaned version[18] of the ISCAT online resource. ISCAT, as mentioned in Sec. 2.2.5, is a repository of image schema sentences taken from a large variety of original sources, mainly from literature (e.g. [104, 159]), but also from some online sources (e.g. MetaNet, newspaper articles), which are listed in the cleaned repository. The sentences from the excerpt were manually annotated with one IS per sentence.

In this first evaluation run, we selected 99 out of 2,478 sample sentences from the cleaned ISCAT excerpt. The main reasons for this extreme reduction of the evaluation set is due to the fact that the sample size per image schema varied considerably and the original dataset only annotates one image schema per sentence. The gold standard is therefore limited to this unique annotation, but image schemas often co-occur in a single sentence or even phrase, and we were interested in whether the image schemas resulting from the evaluation pipeline would at all be possible for the sentence at hand. For that reason, at least for the OpenSesame half of the evaluation, we had to manually analyze all results, so providing a customised manual evaluation in addition to the automated standard evaluation.

To avoid the introduction of bias, further criteria for selecting the set of sample sentences from the larger cleaned repository were (1) variety of original sources, (2) distribution of image schemas, (3) only image schemas already covered in ImageSchemaNet, (4) mixture of concrete (literal) and abstract (metaphoric) examples, (5) English language only, (6) no syntax linearity restrictions, and (7) no minimum or maximum length of the sentence. In terms of variety of sources, we wanted to ensure that not all samples are derived from the same authors, addressing similar ideas or scenarios.

The evaluation setting uses two frame parsers with entirely different architectures, in order to get a finer assessment of the effect of ImageSchemaNet in the process. The parsers include OpenSesame [219] and FRED [98]. OpenSesame is an end-to-end system focused on frame (and semantic role) detection. Its trained model is based on softmax-margin segmental recurrent neural nets. As with most NLP tools, OpenSesame labels extracted

[17] Image Schema Database procured by Jörn Hurtienne
[18] Available here https://github.com/dgromann/ImageSchemaRepository

Table 2.2: Distribution of sentences per image schema

Image Schema	Count
Containment	33
Center_Periphery	19
Source_Path_Goal	17
Part_Whole	14
Blockage	10
Support	6
Total	99

textual segments rather then trying to abstract them as entities and their relations in a knowledge graph. FRED, described in Sec. 1.6.1, is a hybrid knowledge extraction system with a pipeline including both statistical and rule-based components, aimed at producing RDF and OWL knowledge graphs, with embedded entity linking, word-sense disambiguation, and frame/semantic role detection. The big differences between the two parsers are supposed to make evaluation nuances emerge across parsing paradigms (string-centric vs. entity-centric, informal vs. logical representation).

In order to evaluate ImageSchemaNet, we automatically parse natural language sentences in order to annotate them with frames from FrameNet, and we use these frames to get the activated image schemas as encoded in ImageSchemaNet. We then compare the automated annotations to the manual ones, in order to estimate the accuracy of the process, so providing the first results for explainable image-schema detection in natural language texts. Explainability is granted by the heuristic abstraction applied in ImageSchemaNet and in its usage with the parsers.

The image schemas covered in Framester and their frequency in the evaluation dataset are reported in Table 2.2, where we can notice that considerably more examples for Containment were included than for the other image schemas. This distribution was selected to reflect the image schema frequency in the original dataset, with by far fewest examples for Support. Finally, both concrete, i.e., directly relating to a physical or real scenario, and non-physical, i.e., transferring physical aspects to a more abstract scenario, such as MIND AS A CONTAINER, sentences have been selected. The evaluation corpus is available on the ImageSchemaNet GitHub[19].

2.3.2 IS Evaluation procedure

In order to measure the coverage of ImageSchemaNet, an initial trigger frame is required, to evaluate whether that frame leads to the correct image schema profile. To this end, we used natural language sentences as initial frame triggers and implemented a two-step

[19] https://github.com/StenDoipanni/ISAAC/tree/main/ImageSchemaNet

Table 2.3: Comparison of retrieved Image Schemas and Frames by OpenSesame and FRED

Parser	Frame types	Frame tokens	IS Types	IS tokens	SP Types	SP Tokens	IS-Annnotated Sentences
OpenSesame	15	57	6	78	2	11	53 / 99
FRED	43	124	6	126	6	20	75 / 99

pipeline. First, we parse natural language sentences with OpenSesame [318] and FRED [98], which return frames for sentences in the test set. Second, frames are in turn used to query ImageSchemaNet, and identify potentially activated image schema profiles.

To evaluate the final result set from this approach, we first performed automated evaluation (against the original manual IS annotation) applying standard information extraction measures: precision, recall, and weighted F1 score. Each natural language sentence in the evaluation set is annotated with exactly one image schema. However, as shown in Table 2.3, multiple image schemas are often co-activated in individual sentences or even phrases, and ImageSchemaNet enables the detection of more than one IS per sentence. For that reason, we have performed a second manual evaluation process to establish whether any returned image schema(s) is plausible for a given sentence.

2.3.3 Evaluation results

The dataset described in Section 2.3.1, and available on the ImageSchemaNet GitHub[20], was used to test our pipeline approach for correct linking between frames detected in natural language and underlying image schema. In Table 2.3 we present data about IS and SP activation from the selected corpus, noting a better performance from FRED except for the IS type, which was limited by default by the current ImageSchemaNet coverage of six image schemas. In Table 2.4 we present weighted F1 scores for each frame parser as well as for their confusion matrices in Figure 2.5 and Figure 2.6.

Table 2.4 compares the final results of the two parsers on the evaluation set of 99 sentences, where the last column represents the number of sentences which were actually processed. In fact, due to the brevity of sentences, or the metaphorical language or non-linear syntax, in some cases no graph/parsing is produced, resulting in blank result sets, and consequently no image schema detection is possible.

For several sentences, the results of both parsers lead to a co-location of image schemas, i.e., the result set contained more than one image schema, which we counted as correct if the set contained the correct image schema. Please be reminded that multiple image

[20]The dataset is available here: https://github.com/StenDoipanni/ISAAC/blob/main/ImageSchemaNet/FRED_swj_IS_eval.xlsx

Table 2.4: Comparison of weighted F1 scores by parser

Parser	Precision	Recall	Weighted F1	Processed Sentences
OpenSesame	33.95	24.24	26.89	86 / 99
FRED	78.90	39.80	46.06	98 / 99

schemas might be correct for a single sentence, and no inconsistency or incompatibility could come from co-location of different image schemas in the same sentence, even if each sentence in this dataset is manually annotated with only one. For the sentences where the result set of the parser contained multiple image schemas, 11 result sets contained the correct manually annotated image schema when using OpenSesame and 15 result sets when using FRED. The overall results of FRED are significantly higher than those of OpenSesame, due to the fact that the former identifies more frames and synsets activating a correct image schema. This is also reflected by the absolute counts of sentences for which frames activating some IS (at least one) were detected, which are 75 for FRED and only 53 for OpenSesame, as shown in last column of Table 2.3. Table 2.3 shows that FRED individuates 124 occurrences of frames activating some IS, for 43 different frame types, while OpenSesame activates 57 frame occurrences for 15 types. Note that this number is not the total number of evoked frames, but only those that activate some IS. There are 126 IS activation occurrences (tokens) for FRED and 78 for OpenSesame, while both parsers retrieve at least one IS from each of the six types covered in ImageSchemaNet. Spatial Primitives are mainly activated by verb arguments in OpenSesame: 11 tokens for 2 types, both SP of Source_Path_Goal; and by WordNet synsets in FRED: 20 tokens for 6 types, from Containment, Source_Path_Goal, Part_Whole and Center_Periphery. Tables with IS and SP activation by both parsers are available on the ImageSchemaNet Github[21], jointly with the FRED knowledge graphs for all the processed sentences.

A confusion matrix is provided for the results derived from each parser. Fig. 2.5 for OpenSesame and Fig. 2.6 for FRED provide the true labels on the vertical axis and the predicted labels on the horizontal axis. Zero confusion is represented as white space for a clearer visualization and there is no true label for the class *NO_IS* since all sentences in the dataset were annotated with one image schema, however, this class is used to show when no image schema could be returned from the pipeline. When using OpenSesame to detect frames and derive IS (Fig. 2.5), a total of 94 examples are represented, since 5 sentences lead to a result set of more than one image schema that did not contain the true label. *NO_IS* therefore means that no image schema was returned from the detection process (46 sentences with OpenSesame). The most confusing image schema is Containment, possibly since, as shown in Table 2.2, it is the most common in the corpus. For 11 sentences, frames

[21] `https://github.com/StenDoipanni/ISAAC/tree/main/ImageSchemaNet`

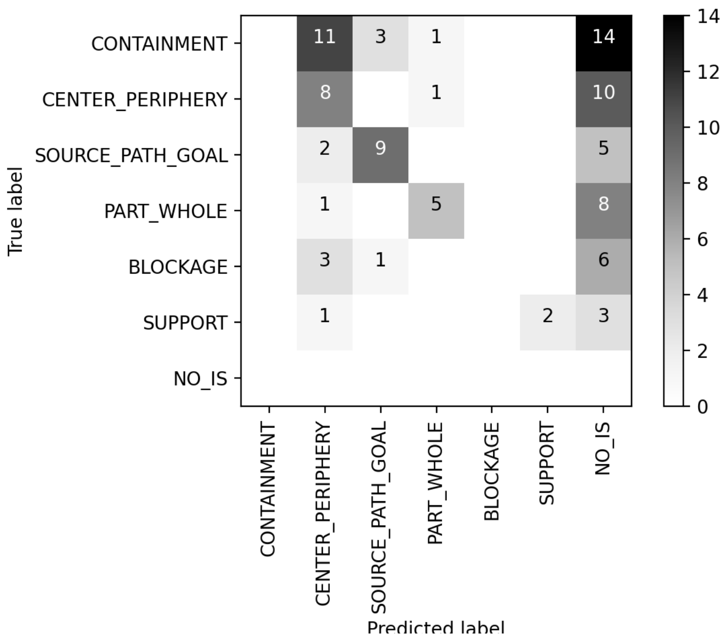


Figure 2.5: Confusion matrix for true and predicted image schema labels using OpenSesame

linking to Center_Periphery were returned, e.g. `Locative_relation` for the sentence *There was passion in her eyes*. Overall, a tendency to confuse other image schemas for Center_Periphery can be observed. A possible explanation for this tendency comes from the correlation between the semantic extension of these broad IS, and their activators in ImageSchemaNet. Apparently, the semantics of Containment, Center_Periphery and Source_Path_Goal overlaps in examples like *He stepped in the middle of a difficult situation*: "difficult situation" can be conceived as a Container for the agent; as an area, in which the agent lies in its Center; or as a destination of the agent as a moving entity along a Source_Path_Goal. This simultaneous activation leads to IS combination (cf. [139]) like Crossing_Opening, Going_in or Going_Through.

When using FRED to detect frames and derive IS, the highest number of correct results is obtained for Source_Path_Goal (Fig. 2.6). However, there was understandable confusion across this image schema and Containment, when frames, such as `Motion`, were returned, e.g. for the sentence *The whole situation spiraled out of control*. This confusion is understandable, since Containment and Source_Path_Goal are frequently collocated as movements in and out of a Container. Spatial primitives are activated by WordNet synsets: 76 tokens for 45 types; and FrameNet frame elements: 15 tokens for 8 types activating Source_Path_Goal and Part_Whole; all of them are core frame elements. Out of 99, only 91 samples are represented in this confusion matrix, because the remaining 8 sentences lead to a result set of more than one image schema that does not contain the true label.

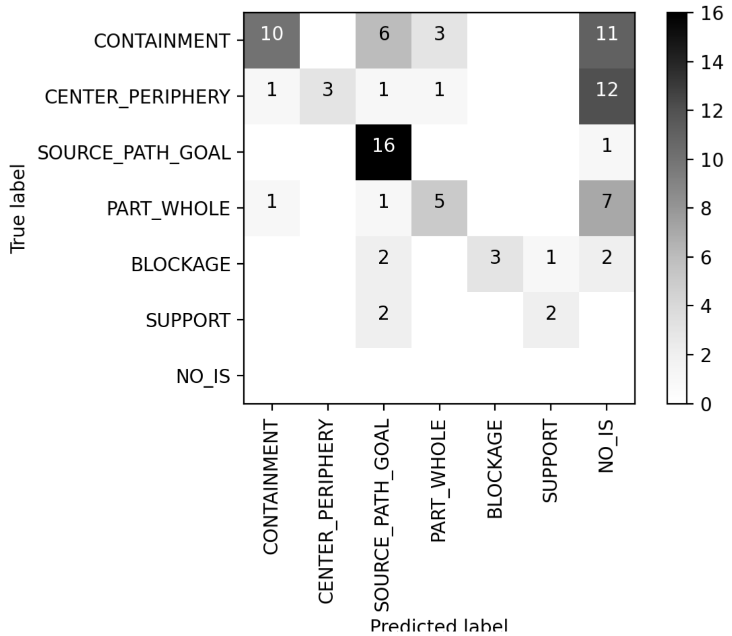


Figure 2.6: Confusion matrix for true and predicted image schema labels using FRED

Even though there is room for improvement, as shown in Table 2.4 these results show that this idea of interlinking frames with an image-schematic layer in Framester is promising. The main bottleneck at the moment is the frame parser. For instance, a strong preposition to frame detection component in FRED could drastically improve these results, since prepositions are currently considered in the tool only as features to detect or generate roles, however, they provide a very strong indicator for spatial language and type of image schema (see also [93]).

The evaluation dataset is available on the ISAAC GitHub[22], while the OpenSesame parsing file, FRED knowledge graphs generated from text, and manual IS and SP detection files can be found at the ImageSchemaNet GitHub[23].

We have manually inspected the returned image schemas with respect to whether (a) the returned image schema that does not correspond to the original gold standard label could be correct, and (b) whether several returned image schemas actually apply to the sentence at hand. For instance for (a), the expression *We are approaching the end of the year* is labeled with CENTER_PERIPHERY, however, clearly shows a collocation with SOURCE_PATH_GOAL. And for (b), for instance, *My symptoms went away* is labeled as CENTER_PERIPHERY. FRED parser, as shown in Figure 2.7, detects three frames: `Motion`, `Travel` and `Departing`. All of them activate SOURCE_PATH_GOAL but `Departing` also activates CENTER_PERIPHERY, which is the label from the ISCAT repository. OpenS-

[22] The ISAAC repository is available here: https://github.com/StenDoipanni/ISAAC
[23] https://github.com/StenDoipanni/ISAAC/tree/main/ImageSchemaNet

esame, on the contrary, as shown in Figure 2.8, detects only `Motion` from the verb *go*, but recognizes the `Motion` frame element `Distance`, which has a `:coreSPActivation` towards Path and Periphery. Consequently, the `:IS_Profile` according both to FRED and OpenSesame shows a co-activation of Source_Path_Goal and Center_Periphery.

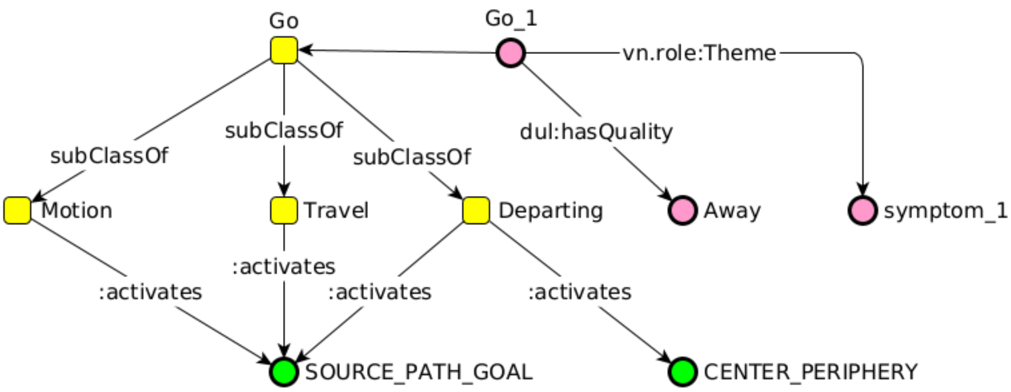


Figure 2.7: FRED graph with image schemas activation for *my symptoms went away*

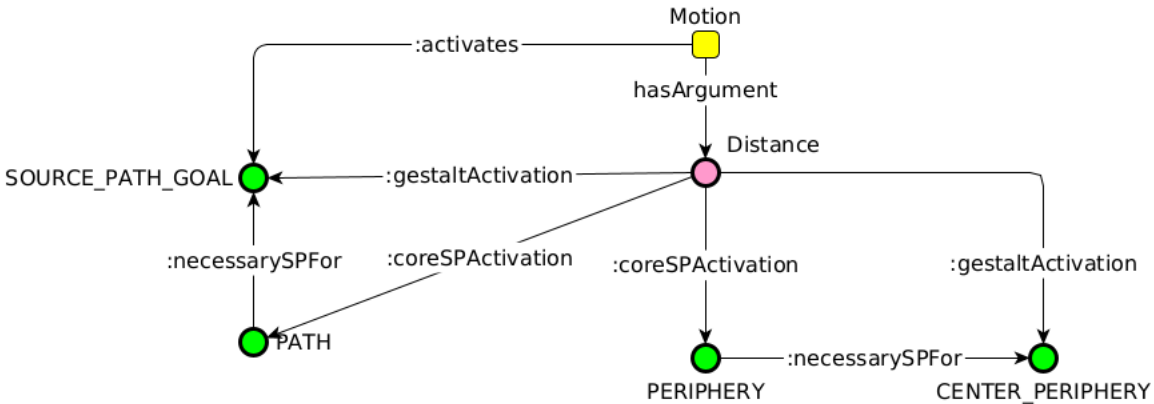


Figure 2.8: OpenSesame graph with image schemas activation for *my symptoms went away*

The method based on OpenSesame showed a preference for Center_Periphery irrespective of the gold standard label. In 40 of 53 annotated sequences the method returned one image schema, of which 13 were correct and 16 returned reasonable image schemas even though not corresponding to the gold standard label. The remaining 13 sequences of 53 were annotated with more than one image schema, out of which 10 contained the gold standard label, and 9 out of the 13 represented correct image schema collocations.

The method based on FRED frequently returned Source_Path_Goal when the label is Containment, where in all of these six cases a collocation of both could be observed, e.g. *Try to get out of those commitments*. Interestingly, slight lexical variations would result in the same set of image schemas, e.g. *He took the problem apart piece by* *piece* and *He tore the problem apart looking for its solution* would both be annotated with Blockage, Source_Path_Goal, and Part_Whole, whereas the gold label only considered Part_Whole. For FRED, 44 natural language sequences out of 65 annotated only resulted in one image schema, of which 24 corresponded to the gold label and 13 where image schemas that can be considered also correct. In 21 cases the method relying on FRED returned more than one label, of which 14 contained the gold standard label and 11 provided reasonable collocations of up to three image schemas.

2.3.4 Discussion

Evaluation results, even from this small corpus, open ample room for discussion. Firstly, some IS, once associated with their lexical grounding, seem to be more intelligible than others. E.g., Part_Whole is the foundational gestalt structure at the core of image schemas, to the point that it can lead to its pervasive detection. Part_Whole is primarily activated by the conceptual frame `PartWhole`, but other frames seem eligible, e.g., `BodyParts`. In this case, the large lexical evocation of `BodyParts` would activate Part_Whole, as described in Section **??**, however, this activation would result in cases like "liver" activating Part_Whole. This might be acceptable, because a liver is a part of the human body, but it might sound less relevant to include being a body part as a core schematic semantics of all sentences about livers. A strategy we implement is, depending on the case, to use as activators both a frame and its lexical grounding, or only the frame, or even selected synsets only. This is possible thanks to the ImageSchemaNet object property diversification.

Another example for discussion is Containment. Some abstract concepts and doctrines, e.g., *humanism*, evoke the `Containing` frame, hence activating Containment IS from ImageSchemaNet. In this experiment, we have followed the FrameNet-based heuristics, and indeed such abstract concepts might be likely used as Container s in many contexts. This hypothesis could be tested on a larger lexical corpus, including also longer texts, and lexical data can be analyzed in their different aspects, considering also e.g., their semantic type.

Referring to the initial example "water is sold in bottles", an autonomous agent operating in uncertain conditions being able to make inferences starting from the semantic type of an entity, could be able to make, starting from the lexical unit, the inference that, if the WordNet synset `water-noun-1` has a `:semTypeActivation` of substance, then, in order to be moved on purpose and in its integrity, it is necessarily contained in some Container. In this specific case, the `waterbottle-noun-1` synset has a `:lexicalSenseActivation` to Container.

Finally, some activations are intrinsic to the commonsense semantics of a frame or lexical unit. It is the case of the `Storing` and `Ingestion` frames, which in three occasions are the sole elements that allow the correct detection of Containment. However, in the sentence *He tore the problem apart looking for its solution* we face a false positive, since the correct activation of Part_Whole is not due to "tear" or "apart" or a combination of both, but stems from a wrong disambiguation of "solution" in a chemistry-related sense of a compound of particles.

2.4 A Long Path

In this Chapter we presented the ISAAC ontology, its operationalisation ImageSchemaNet, and experiments related to automatic detection and extraction of image schematic knowledge from natural language.

The final resource consists of more than 40,000 triples, which formalizes image schemas with Framester semantics, providing an image-schematic layer to FrameNet, MetaNet, WordNet, VerbNet, and other resources in the Framester hub. ImageSchemaNet has been built starting from image schema definitions and examples in literature, and provides lexical grounding (i.e., lexical activators) for the detection of image schemas or spatial primitives. Activation can be retrieved via SPARQL queries on the Framester hub. ImageSchemaNet allows non-trivial image schema profile extraction from various semantic layers, including disambiguated natural language units from multiple semantic resources, semantic roles, frames, semantic types, and individual entities. This extraction has been exemplified in an empirical evaluation of annotating natural language sentences with frame parsers and ImageSchemaNet.

The extensions of ISAAC can be related to the formalisation of IS themselves e.g. the introduction of still lacking ones like Verticality, Scale, etc. As a direct consequence, other than a quantitative improvement of the resource, this extension would enable further investigation on relations among image schemas, in order to clarify possible taxonomic, lexical, functional, mereological and usage relations between IS, bringing greater clarity on frame compositionality and the related underlying commonsense reasoning. Furthermore, ISAAC aims at being a multi-modal ontology, and its ontological form allows the integration of several types of data, to mention a few of them: (i) new textual annotation improving the ground truth with further corpora other than the ISCAT repository; (ii) visual data, including image-schematic annotation of spatial relation in visual data (note that for repositories such as VisualGenome [181], it is already possible to extract image schematic knowledge thanks to the annotation with WordNet synsets); (iii) numerical data about electric neural impulses, revealing path activation for fMRI-like experiments related to the neural grounding experiments focused on one or more image schemas. From an operational point of view, the image-schematic extraction could be improved considering prepositions in the parsing process, which are currently underexploited. We also envisage to integrate recently proposed BERT-based frame detection algorithms (e.g. [322]).

Chapter 3

Moral, Cultural and Individual Values

This chapter is focused on values as intended in ethics, the formalisation of theoretical frameworks and the operationalisation via developing an automatic value-detector.
In Section 3.1 several works are mentioned from different disciplines, including general definitions and descriptions. The final paragraphs expose different approaches formalising the notion of *value*, and methodologies of extraction of moral content from text. In Section 3.2 the ValueNet modular ontology is presented. Sec. 3.2.4 describes the ontological transposition of Moral Foundations Theory, Sec. 3.2.2 does the same for the Basic Human Values theory, and Sec. 3.2.7 for Moral Molecules. The detailed description of aspects related to specific values, their taxonomy, the relations possible among them, and clarifications about their ontological status is deferred to the specific section focused on the ontological module formalising the theoretical fragment considered.

3.1 The Meaning of Value

Probably due to the semantic pervasiveness of the notion of values, perceived as importance, worth, or utility, they have been the object of several disciplines, such as economics, philosophy, sociology, anthropology, psychology, and recently computer science and AI. Given the vastness of disciplines and approaches, and the temporal extension of the research that has been carried out on the notion and importance of values, this section provides an overview of the main areas and references, followed by a focus on the theories taken into consideration in this work. The perspective adopted in this work is the cognitive and social psychology one, which is in turn grounded in philosophy, and is developed in parallel with the socio-economic one. It is possible to individuate, in the history of philosophy, at least five major ethics movements that deal with the notion of value:

Virtue ethics These models, developed by philosophers such as Plato, Aristotle, and Confucius, focus on the individual performing some action, rather than on the action *per se*. According to this model, a morally praiseworthy person would make good moral decisions, and the goal of ethical education is to cultivate virtues such as wisdom, courage, justice, compassion, and moderation. It is traditionally focused on achieving *eudaimonia*, a Greek term commonly translated as "happiness" or "welfare". More recently, philosophers like Rosalind Hursthouse and works like Nussbaum's Capabilities theory [235], or MacIntyre's *After Virtue* [208] situate neo-Aristotelian thought in the globalisation socio-technical environment;

Utilitarianism This model is a consequentialist one, meaning that the definition of some action or process' morality, depends on the outcome and consequences caused by that action/process. It was originally developed by philosophers such as Jeremy Bentham and John Stuart Mill [18]. It proposes a family of normative theories, according to which, the best action is the one that maximizes overall happiness for the greatest number of people. This model is based on the idea of maximizing utility, or the overall value or benefit of an action.

Kantian ethics This model, developed by philosopher Immanuel Kant [165], gives birth to Deontology, and it proposes that the moral worth of an action is determined by its intention, rather than the consequences of the action. According to this model, an action is moral if it is *a priori* motivated by a sense of duty or respect for moral laws, rather than by an individual utility such as personal gain or pleasure.

Natural law theory This model, developed originally by Thomas Aquinas [4], proposes a universalistic view, for which there is a unique moral law that is inherent in human nature. This law can be discovered through reason. According to this model, moral laws are based on the nature of entities and are independent of human opinions or cultural variables. More recently, Finnis [76] individuates 7 "basic goods" that contribute to a fulfilling life: "Life, Knowledge, Play, Aesthetic Experience, Sociability, Practical reasonableness and Religion";

Social contract theory This theory suggests that the social and political order is based on an agreement (the social contract) among individuals, and this agreement is guided by certain moral and ethical principles. Several versions of social contract theory are grounded in the idea of natural rights, such as the right to life, liberty, and property, inherent and inalienable, to all individuals. These are considered to be fundamental values that should be protected by the institution. Moral Molecules theory [42], described in Sec.

3.2.7 is a development of social contract theory. Its fundamental assumption is that morality is rooted in the agreements and contracts that people enter with one another to live and cooperate in society. Floridi's concerns [324] about the threshold of privacy given away in exchange for security can be considered a part of this last current.

3.1.1 Morality vs Values

Although it seems counter-intuitive, the study of values and moral ethics are not considered a unicum, but rather two domains that have long developed independently in separate directions and with very different intentions. Vauclair [335] exposes the issue of the very notion of "moral value": it collapses moral ethics studies with the distinct research tradition of value theories.

Moral psychology is traditionally not interested in the specificity of "values", but in a more broad attitude towards "what is good" or "what is right" while social psychology, focused instead on values, is not concerned in investigating the notion of morality.

Morality in Moral Psichology Moral cognitivism descending from Kohlberg, in line with Kantian and deontic ethics, even refuses moral values [179] as an impactful phenomenon on morality. The focus of these studies is a kind of formal morality, and it intends to axiomatically define the various stages leading from childhood, taking up Piaget's learning theories [245], to moral maturity. Moral maturity should be based on logical reasoning, referring to a Natural-Law-ethics-oriented view, and focus on a universally shared notion of rightness, fairness, and justice. Recent moral psychology works still use forms of this "moral development ladder" [329].

Morality in Social Psichology Gilligan and Turiel criticized [107] Kohlberg's approach for not separating social conventions from moral rules, and introduce the Social Domain Theory, in which three domains are recognised: (i) the moral, (ii) the societal conventional, and (iii) the personal [234]. The (i) moral domain is focused on how people should behave one in relation to another. The (ii) societal domain is related to regulations promoting the successful collaboration of groups and social/institutional architectures. The (iii) individual domain focuses on psychological dimension: understanding of self and self-other relation, as well as beliefs about dominance, individuality and autonomy. Nevertheless, these three domains have been demonstrated to be blurred and cross-culturally unstable [180].

Being willing to move some steps toward a cognitive and morally aware AI system implies adopting a socio-cultural view of morality [323] in line with Campbell et al.'s [34] view of moral psychology, intended as focused on the mere understanding of values that human beings acquire, independently of the morality or immorality of their ideology:

> Acknowledging a plurality of moral conceptions does not mean endorsing all as equally right or good, [...it means instead] to explain how human beings could arrive at such diverse endpoints through the normal process of development

To which the philosopher Louis Pojman adds that 'values are central to the domain of morality. [And from] our values, we derive principles.' [251].

Henry & Reyna [140] defines morality and its relation with culture as follows:

> When a person or group behaves in ways that are consistent with important cultural ways, they demonstrate that they are upholding the moral foundations of the society and contributing to the overall benefit of its members [...] On the other hand, those who violate values threaten the stability and moral foundations of that society and are therefore likely to be shunned or ostracized from that group.

As a follow-up to this, the morality evaluation process has been divided into two subprocesses: (i) the explicit process, based on logical reasoning, more similar to a Natural Law approach, and (ii) the implicit/tacit process. The second one operates at a lower, unconscious level. It is automatic and relatively effortless, in Kahneman's terms [162] the explicit process is slow thinking, while the tacit process is fast thinking. This bipartition has been investigated [194] in particular in the context of Haidt's Social Intuitionism Theory [125]. Several of Haidt's works propose that human beings base their moral appraisal mainly on "gut feelings" than on logical reasoning, to discern right from wrong. This human ability is called "moral intuition" and it is considered both innate and learned or enculturated. Haidt, furthermore, asserts that moral intuitions come first, and proposes a causal relationship linking intuition and moral judgment. Moral reasoning would be a structure to rationally (logically) motivate the original intuition. Literature findings from diverse disciplines such as social psychology, evolutionary psychology, and anthropology seem to confirm Social Intuitionism Theory [125].

Empirical findings have confirmed the importance of affective reactions in moral judgments [129].

Nevertheless, Haidt's position includes also aspects of morality that are said to be universal. In line with evolutionary theorists [29] he suggests that morality evolved universally in the same way because it is the most effective way to face difficulties and gain mutual benefits from successful cooperative relations. Sec. 3.2.7 describes in detail Curry's theory, grounded in the "morality as an evolutionary strategy" assumption. At the same time, cultural differences, therefore a moral relativism dependent on sociocultural and spatial location variables, exist because people from different cultures face different adaptive challenges in different ways and because social (local) norms emerge from social and cultural dynamics peculiar to each and any biological evolution environment. For this reason Haidt's social intuitionism can be said to be effective in its being highly flexible, since, contrarily to more strict Kantian ethics approaches, there are no a priori assumptions of what counts as morally relevant.

Ultimately, Haidt's Moral Foundations, described in detail in Sec. 3.2.4, provide basic universal dimensions to ground moral judgments, intuitions, and concerns. The authors suggest using Moral Foundation Theory if one's goal is to study the sources of moral intuitions across cultures [111]. If one's goal is instead to describe moral discourse across cultures, then Shweder et al.'s [210] three ethics (autonomy, community, and divinity) could be used, while finally if one is interested in examining specific values, then Schwartz's value types [288] are the best choice.

3.1.2 Value: a Multifaceted Concept

Given this introduction, we move now to the notion of value itself. The term "value" has been used to refer to psychological constructs such as attitudes, beliefs, and norms [142] and more broadly to interests, pleasant state of things, preferences, desires, wants, goals, as well as, in their jussive semantic shade, duties, moral obligations, needs, and individual orientations in general [272].

We provide here several definitions that have been used in different domains but to give an idea of the vagueness surrounding the subject matter, consider that in a 1980s census, values definitions amounted to 180 different conceptualizations [132].

According to the Oxford Dictionary values are 'beliefs about what is right and wrong and what is important in life', and, if we exclude the sense related to economics, also 'the quality of being useful or important' [61].

In the Encyclopedia of Sociology, values are defined as: 'Evaluative beliefs that synthesize affective and cognitive elements to orient people to the world in which they live' [217]. Rokeach, pioneering and considerably innovating research in the value domain, [273, 274] defined a value as 'an enduring belief that a specific mode of conduct or end-state of existence is personally or socially preferable to an opposite or converse mode of conduct or end-state of existence'. For Smith and Schwartz, cultural values are commonly conceived as socially shared, abstract ideas about what is good, right, and desirable in a society or cultural environment [20].

Values are divided according to Rokeach **??** into two types: instrumental values, i.e. prescriptive "modes of conduct" and telic values i.e. end-goals to be pursued (e.g. freedom, equality, or sanctity [304]). A personal moral value has been defined by Scott as 'a particular individual's concept of an ideal state of affairs or relations among people which he uses to assess the 'goodness' or 'badness'; the 'rightness' or 'wrongness' of [what] he observes' [298, 179],

This proliferation of definitions with segmented margins is explained by Hechter [134] because of the four following impediments:

1. Values are unobservable;
2. Value theories give little guidance for understanding how values shape behavior;
3. Behavioral explanations are unconvincing since the process that generates values is unknown;
4. There are difficult problems with measuring values.

Although 30 years have passed since the formulation of these four pitfalls, it seems that they are still solidly valid today, and nevertheless, values, as intended in ethics, are part of the commonsense socio-cultural norms that shape the 'general frame of reference for living' [209] and are described by Vine as 'normatively prescriptive constructs by which we evaluate and regulate our social actions and shared lives' **??**.

Economic Value In the field of Economics "value" is intended as the 'relative worth of a good or service, given a certain price' and it is usually related to the idea of subjective utility deriving from it. Albeit it is not the sense that it is intended to be covered in this work, it is worth mentioning that it is a specification developed by Adam Smith [305] as the necessary amount of labor to produce some good. In particular, the aspect of subjective utility is instead attributed to Marshall's *Principles of Economics* [218]. It is furthermore developed by Keynes [169], the father of modern macroeconomics, with a shift in perspective, introducing the notion of "expectation", for which value becomes the amount of expected (future) returns from a good or service.

Value as Cultural Actor In Amartya Sen's [299] treatment, the concept of value makes a further disciplinary leap, as it becomes the relative worth of a good or service measured by the possibility of a person reaching the aforementioned state of affairs. But the notion of value as a "desirable state of being" had been already introduced in 1947 by Parson [241] and Weber [344], as a sociological variable to be considered determinant to understanding the shapes and mechanisms of any society's economic system. Values begin to be studied as active cultural actors and binding social ties in the 1980s, in particular following the works of Bourdieu [28], and contemporary to it, from a more cognitive psychology perspective, Kahneman and Tversky's analysis on the role of values in decision-making scenarios [163]. Values are, in fact, relevant (if not determinant) in our behaviour and everyday decision-making, designing boundaries for our conscious self by framing knowledge about what we *should* and what we *desire* [274], [288], [296], [233].

Value, Desire and Goals Values, in their broadest definition, points at anything wanted/desired by a person [171]. This dimension of desirability is a key point in the search for an explicative definition of the ontological nature, i.e., understanding the network of relations that the concept of value weaves with the cognitive and factual circum-stratum. Kluckhohn [177] underlines that it is critical to distinguish two dimensions in the semantics of values: (i) the "desired" and (ii) the "desirable". The notion of desire seems in fact to be common to all cultures, as well as ways to express it. However, this is not the same as "desirable": while "desire" expresses an individual preference, or a bias towards an alternative in a real or hypothetical scenario, "desirable" conveys the idea of a standard that is felt to be morally motivated. The entire social dimension would be impossible without values as the "desirable" [176] and regarding this dimension, Smith asserts that 'values pertain to the desirable, the preferable, rather than the merely desired or preferred; to the realm of 'ought' rather than that of [...] 'want'.' [307].

Value as pivotal goal-pursuing motivation has been largely investigated by Regulatory Focus Theory, and in particular by Higgins **??** underpinning the idea, supported as mentioned before by empirical evidence too, that the emotional and value layers are not distinct, but they are two co-occurring systems in their manifestations, binding the individual to the environment to which he or she is part, enacting an internalisation of socio-cultural norms whose violation or observance impacts the emotional system, with a positive reinforcement towards cooperation and collaboration [55, 180] and a negative one towards transgression [1, 141, 13, 157].

Value Core Bilsky and Schwartz, when investigating the semantic extension of values [25], describe them as structures parallel to social norms, with two main differences: (i) they are not explicit in their rules or formalisation, and (ii) their sanction-reward system operates on the emotional layer [277]. In particular, they highlight five prototypical features:

- They are considered and treated as concepts or beliefs;
- They inhere some desirable state of the world or behaviour;
- They can occur in specific situations, but they transcend them;

- They are pivotal for selection or evaluation processes;
- They are often organized by relative importance.

The last point, in particular, is the one mentioned by studies assuming the necessary scalar nature of values [332].

Values are therefore inextricably related to the individual perception, background commonsense knowledge, culture, and perspectivisation [31], as well as any online or offline interaction, expression of personal positions, epistemic stance, and freedom of judgment, although the perspective of values differs from deontic reasoning since, in van Fraassen's words [332], *deontology*, or the theory of obligations 'deals with what ought to be because it is required by one's station and its duties, by the web of obligations and commitments the past has spun', while, social obligations, intended as Kantian schemata, are a product of the human reason and they are time and space contextually dependent. *Axiology*, or the theory of values, 'deals with what ought to be because being so would be good, or at least better, than its alternative'.

Moral Emotions Nevertheless, social norms are expectations of 'how people are supposed to act, think, or feel in specific situations' [253]. These can be in turn distinguished into injunctive norms, namely, determining positive or negative appraisal for (the majority of) others, and descriptive norms, namely, what is *usually* done or considered in a span of social standards [38]. Cialdini and colleagues [38] have demonstrated that injunctive norms and moral values share the concept of how something *ought* to be, and of moral (and social) approval. However, values and norms differ for the fact that social norms always refer to a prescriptive behaviour, while this is not true, as seen before, for telic values. Morality can be regarded as one of the culture's rule systems **??**.

Furthermore, values are particularly relevant in the dynamics of appraisal and decision-making tasks [314], since our choices and preferences are typically affected by our values [274, 25], and by the emotions arising from value-driven appraisal dynamics [258].

In social psychology, the Contempt-Anger-Disgust (CAD) triad model of moral emotions proposed by Rozin et al. [277] relates these to specific configurations of values, termed "ethics", inspired by Schweder's work [210] on morality from an anthropological perspective. The CAD triad model relates each emotion type to the violation of a specific ethic: Contempt to the Ethics of community, Anger to the Ethics of autonomy, and Disgust to the Ethics of Divinity. These ethics can also be seen as a subset of the value-violation dyadic opposition (e.g. Care vs Harm) constituting the Moral Foundations Theory put forth by Haidt and colleagues [127].

Value as Neuro-cognitive Embodied Phenomenon Finally, from a neuro-biological perspective [45]:

> There was a biological blueprinting for the intelligent construction of human values [...] We also believe that a variety of natural modes of biological responses, which include those known as emotions, already embody such values.

Damasio and Gallese [46, 47, 82, 85] theorize the interdependence of the emotional layer and moral one, as well as their activation of specific brain areas and treat them as neuro-cognitive embodied phenomena.

3.1.3 Values Frameworks

The frameworks that are being covered in detail in this work, and which correspond to transposition as an ontological module, are the Basic Humans Values (BHV) theory, by Shalom Schwartz [288], and the Moral Foundations Theory (MFT), by Graham and Haidt [110].

These two models were chosen for three main reasons:

1. They are pillars of the current state of the art of cognitive psychology that seeks to include aspects of embodied cognition, social psychology, personal perceptual dimension and the emotional knowledge layer;

2. They have a universal dimension: they aspire to explain personal internalisation and occurrences of manifestation of moral behaviours, set within a framework of values shared cross-culturally, by the fact of being human beings;

3. For the two reasons above, and because of the resonance the discipline has had with other social sciences and with computational applications, both theories have been extensively tested through questionnaires, and used as a reference model for the development of computational techniques for (i) extracting latent moral content from natural language, and (ii) classification tasks, labeling sentences as containing one (or more) values from a specific theory.

Since, as mentioned, BHV and MFT are the main models used as theoretical grounding for classification tasks and value detection an in-depth discussion is provided in Sections 3.2.2 and 3.2.4, where each theory is described, motivating modeling choices for their formal semantics transposition.

Finally, in Sec. 3.2.7 a further theory is transposed into ontology: the Moral Molecules theory [42], a combinatorial system recently introduced by Curry, but, at the current state of writing, it is not operationalised.

The following Section 3.1.4 describes traditional methods based on questionnaires for value extraction, as well as the main computational techniques used to detect values from natural language. Section 3.1.7 is instead dedicated to works focused on formalisation of values, including different approaches and scopes.

3.1.4 Values Questionnaire

Despite the widespread recognition of values' importance, there are only a few empirical studies and surveys operationalising moral values extraction [328, 132].

We describe here the most relevant surveys used for traditional value extraction. For an extended history of value extraction with surveys refer to Pittel and Mendelsohn [246].

Hofstede Hofstede's questionnaire [144] focuses on comparing societies via the analysis of cultural values, measured on four dimension, described as follows: (i) Power Distance: stability of hierarchies, authority recognition behavior, and acceptance of established social structure; (ii) Individualism vs Collectivism: individualistic attitude vs universalistic one, loyalty to in-groups (other than family/blood relations); (iii) Uncertainty Avoidance: acceptance of difference, openness to change vs only-one-truth attitude; (iv) Masculinity vs Femininity: achievement, heroism, hawkish behavior vs cooperation, modesty, dovish behavior and improvement of quality of life. Using information from IBM employees, he described the value profiles of 53 nations or cultural areas. Several studies have expanded on Hofstede's conclusions (e.g. [161]). However, this measure is not meant to be used to correlate people's value orientations with their beliefs or actions. The metrics it uses, such as individualism and power distance, discriminate between country cultures but not between specific people. Additionally, the majority of the Hofstede items discuss work ideals. The variety of human values that are important in many spheres of life is not measured.

Rokeach In the Rokeach questionnaire [274], respondents are asked to rank two sets of values: 18 "terminal (telic) values" and 18 "instrumental values". The ranking order is from the most to the least significant for the respondents, considering them as "guiding principles in their lives".

According to Braithwaite Scott in [270], numerous investigations using this scale have found relevant connections between values and a range of demographic factors, beliefs, attitudes, and behavior. Despite attempting to cover every aspect of human values, this scale omits important information (e.g., tradition and power values). Predictions and explanations based on the selection of objects are often ad hoc because they were not theory-driven.

Inglehart Both their four and twelve-item versions of the popular Inglehart materialism/postmaterialism measures (MPM) have a theoretical foundation, respondents in representative samples appear to understand them, and they have relevant relationships to numerous variables of interest to survey researchers [153]. This scale only indirectly gauges people's values. Instead of asking about personal aspirations, it asks about choices among potential national goals. These preferences probably reflect how much a person values freedom, self-expression, economic and physical security, as well as the general quality of life. However, responses to inquiries concerning one's political, economic, and security objectives for their country are likely to be influenced by a variety of personal beliefs. For instance, selecting "preserving freedom of speech" as society's top priority likely reflects personal ideals of tolerance and intellectual openness. However, depending on the specifics of their personal or societal situation, an intolerant conservative fringe member who is afraid of governmental intrusion might make this decision. Furthermore, only one value dimension is measured by the Inglehart scale. Despite being broad and significant, the MPM dimension is not precise enough to reflect the wide range of personal value orientations. According to research conducted in seven different nations, MPM integrates a variety of distinct value emphases into a single score.

Schwartz Value Survey The most popular tool now used by social and cross-cultural psychologists to explore individual differences in values is the Schwartz Value Survey (SVS). The Schwartz Value Survey assesses individuals' 56 distinct personal values (in its 2007 [290] version 58 assessing ten motivational goals) by asking respondents to rate telic and instrumental values according to being "guiding principles in your life" (e.g., "social justice" intended as rectifying injustice, caring for the weak)) on this scale. Ten theory-based value orientations are measured by these particular values. The distinctiveness of these value orientations has been supported by studies in over 65 nations [289]. Schwartz colleagues [292] define values as 'desirable, trans-situational goals, varying in importance, that serve as guiding principles in people's lives'. He derives cultural values via the statistical aggregation of individual personal values claimed to reflect shared conceptions of what is good and desirable in some culture [295].

Portrait Value Questionnaire The ten fundamental value orientations tracked by the Schwartz Value Survey are also covered by the PVQ. The PVQ offers concise verbal portraits of many personalities. Each portrait outlines a person's objectives, aims, or desires and subtly emphasizes the significance of a certain value type. The verbal portraits of each person capture the person's values without specifically naming values as the subject of the inquiry by portraying each person in terms of what is important to him or her—the aims and wishes he or she pursues.

Moral Foundations Questionnaire Based on Moral Foundations Theory, construct measurement typically takes the form of the Moral Foundations Questionnaire (MFQ)[111]. This scale has 32 items, but only 30 of them are assessed because of two "catch" questions meant to catch those who react randomly or inattentively. Respondents describe how important a group of issues are to them when making a moral choice in the first part. A variety of moral statements are asked to be rated by respondents in the second section of the MFQ. The replies to each of the six questions for each moral foundation are averaged to determine the score.

World Value Survey The World Value Survey [124] is a wide international research program devoted to the scientific and academic study of social, political, economic, religious and cultural values. Project's overall aim is to analyze people's values, beliefs and norms in a comparative cross-national and over-time perspective involving in their survey people from more than 64 countries. Current version of the questionnaire (WVS 7) consists of 290 questions along 14 thematic subsections, such as "social values, attitudes and stereotypes", "migration", "ethical values and norms", etc. Each question can be answered on a scale usually considering (at least) 4 levels of agreement: "totally disagree, "partially disagree", "partially agree", "totally agree" (or equivalent forms). World Value Survey is currently transposed in ontological form and available in the ValueNet repository[1], but not yet operationalised for further knowledge inferences.

3.1.5 Values Resources

We provide here a description of the main resources and datasets labeled with moral values, or the main vocabularies developed to operationalise a specific theory.

Moral Foundations Twitter Corpus Developed by Hoover et al. [146], the Moral Foundations Twitter Corpus (MFTC) consists of 35.108 tweets. It has been manually annotated according to the dyadic structure of the MFT, for ten types of moral attitudes. The dataset is focused on seven distinct, socially relevant discourse topics: All Lives Matter (ALM), Black Lives Matter (BLM), the Baltimore protests, the 2016 Presidential election, hate speech and offensive language [49], Hurricane Sandy, and the MeToo movement. Each tweet is labeled by three to eight different annotators trained to detect and categorize texts following the guidelines outlined by Moral Foundations Theory. The MFTC includes ten

[1] The World Value Survey ontological module is available here: https://github.com/StenDoipanni/ValueNet/blob/main/wvs.owl

different moral value categories, as well as a label for textual material that does not evoke a morally meaningful response.

Moral Foundations Reddit Corpus It was developed by Trager et al. [327] to provide an alternative to the MFTC with a slightly different type of data (social comments instead of tweets). The Moral Foundations Reddit Corpus (MFRC) is a collection of 16.123 Reddit comments that have been curated from 12 distinct subreddits, hand-annotated by at least three trained annotators for 8 dimensions of the MFT dyads based on the MFT framework.

SemEval 2023 Corpus In the context of the SemEval 2023 challenge the BHV model has been proposed as a reference model to perform automatic classification tasks of strings of text. At the moment of writing the challenge is still in progress, data can be found on the dedicated repository[2].

Moral Foundations Dictionary The Moral Foundation Dictionary (MFD) in its first version [112] is a dictionary realised via manual search from thesauri and annotation of words related to MFT values and violations [111]. It has been extended to a few hundred lexical units per value/violation by Frimer et al. [79] and used as the basis to develop the Extended Moral Foundations Dictionary (eMFD) by Hopp et al. [148].

3.1.6 Value Detection

Extracting moral content from natural language is not an easy task. The difficulties in identifying data with latent moral content, as well as cultural dependence, political orientation, and the inherent subjectivity of the annotation work, make this an especially tough undertaking.

We list here the main methodologies that have proven to be successful in recent years. In the discussion of this section, no detail of the various approaches is given as they are radically different from the one used in this work, and therefore not directly comparable. In Chapter 5, however, we envision the possibility of combining the graph-based methodology proposed here with the current state-of-the-art NLP, DNN, and BERT-based techniques.

Liu et al. [204] adopt NLP techniques to generate a psychographic lexicon of values from the Big Five Factor [263] and the Value Portrait Questionnaire, then inject this lexicon in a DNN recommender system with the purpose to extract values polarity from users' e-commerce reviews.

[2]The task description and data are available here: `https://touche.webis.de/semeval23/touche23-web/index.html`

Lin et al. [200] estimate moral values from tweets by combining textual feature extraction and contextual knowledge from Wikipedia.

Hoover et al. [145] focus on the detection of moral values expression via a Distributed Dictionary Representation in tweets about money donations received after Hurricane Sandy.

The development of ad hoc new value lexica is a technique adopted by Araque et al. [5], and Ponizovskiy et al. [252] to perform value detection from text, while Hopp et al. [148] focus on the improvement of existing resources, developing the Extended Moral Foundation Dictionary. Liscio et al. [203, 202] use hybrid (human and AI) methodologies to identify context-specific values to certain domains like “sustainable energy” and others. Existing previous work on detecting moral values using MFT as the theory of reference from texts can be grouped into two macro areas depending on the methodological approaches: unsupervised vs supervised methods. Moral detection from text is usually based on word count [80] or focuses on features determined by words and sequences [168]. Unsupervised methods rely on models not supported by any external framing annotation. This approach includes also architectures based on the Frame Axis technique [186], such as [228] and [256]. As a point of reference, the main resource used is the extended version of the Moral Foundation Dictionary (MFD) [148], which consists of lexical units concerning the positive and negative values of the five dyads of MFT and a sixth dimension relating terms of general morality. Kobbe et al. [178] aim to link MFD entries to WordNet to extend and disambiguate the lexicon. Recently have emerged works focused on moral knowledge transfer about MFT value dyads applied to different domains [202], and works improving the available corpora (other than the Moral Foundations Twitter Corpus [146]) to test moral values detection tools, like the Moral Foundations Reddit Corpus [327].

Several works [111, 40, 325] use the Moral Foundation Dictionary to develop supervised classifiers, while other approaches exploit Latent Semantic Analysis [57] to generate word vectors [58, 166]. Garten et al. [100], Xie et al. [348], and Nokhiz and Li [232] develop models based on unsupervised learning using GloVe [242] and Word2Vec [224] to embed words and derive a vector representation for each moral value.

A knowledge-based approach is developed by Li et al. [197], but, while this methodology improves the name tagging performance, relation extraction, and event triggering / argument labeling, data about moral values is associated via alignment to GeoNames [346] database, reusing value distribution according to European Social Survey [160] data.

The few works that share a similar approach to that adopted in this work are listed. Rezapour et al. [267] improve the MFD lexicon using WordNet and manual annotation, but yet the moral annotation is by word and not by entity on a knowledge graph.

Some similarities to our method can be retrieved in Lin et al. [200], who use Wikipedia abstracts and DBpedia relations to classify texts based on their moral values.

Finally, Hulpus et al. [149] map MFD to a knowledge graph, and then manually score the relevance of concepts for MFT moral values. They align their work to WordNet 3.1 synsets, ConceptNet, and DBpedia entities. We provide here a table of the results obtained in Hulpus et al.'s work, to allow a comparison with results from our approach, described in the next Section in Table 3.2.

3.1.7 Values Formalisation

From an ontological perspective to the best of our knowledge, there are no attempts to transpose in ontological form the domain of values but one case. We expose here in detail the classes and axioms used in this case, and we motivate why this work was not considered and reused in the ValueNet ontology.

The Ethics Ontology The Ethics Ontology[3] is a project developed by Michael De Bellis in 2018 [56]. It is proposed as a "formal model of a Universal Moral Grammar". Universal Moral Grammar (UMG) [223] is a framework that proposes to formalise morality 'as Chomskian generative grammar did for language'. Albeit the framework per se seems sound, criticism has been raised [65] mainly about two issues: the first (i) is whether or not moral competence is modular, and the second (ii) is whether it has a hierarchical grammatical structure under the hood. These questions still await a comprehensive answer by the current state of the art in the cognitive science, social psychology, and neuroscience literature, which makes UMG an interesting framework that, apart from the Ethics Ontology, has been no further developed or operationalised in a real-world value detection scenario.

The Ethics Ontology is the first (and only) attempt to develop coverage of what UMG would aim at defining. The ontological model is developed in OWL using SWRL rules o inject FOL expressivity maintaining the possibility to make automatic inferences. It also includes the formalisation of elements from the Theory of Mind, integrated into various concepts from developmental psychology and philosophy ethics domain.

We use here the prefix `eo:` to refer to Ethics Ontology in this section. Referring back to Section 1.4 and, in particular, to Fig. 1.3, the first concern in reusing this interesting work is its absence of alignment to any foundational or upper layer ontology. Introducing classes such as `eo:GroupAgent` or `eo:DivineAgent` which deal with the ontological and mereological type of modeled entities could be problematic reusing an ontology with misaligned foundational assumptions. The second main concern is in its being a *com-*

[3] The Ethics Ontology is available online via Web Protégé here: https://tinyurl.com/EthicsOntology-3-2-18

pendium of theoretical fragments without clear provenance. To take an example relevant to the ontological modeling developed in this work: the `eo:MoralValue` class is modeled as superclass of several classes, among them: `eo:Evil`, `eo:Fair` `eo:Utility`, and `eo:Truth`. The first problem in this modeling is that the lack of annotations risks generating doubts about the theoretical soundness of these assertions. Assuming therefore that this is a theory introduces as new, then subclasses such as `eo:Evil` and `eo:Fair` seems to be plausibly aligned to values from existing theory such as `mft:Harm` and `mft:Fairness`. But `eo:Utility` seems ontologically a different type of entity from the above-mentioned, it could be, considering the notion of "moral value" as a frame, one of its semantic roles, which could be modeled then as "individual utility" and "societal utility", but utility per se seems hardly classifiable as a value. A further problem is the notation used. For some entities such as `eo:Truth`, it is used as a noun, while for `eo:Fair`, it is used as an adjective. The lack of further annotation does not allow us to distinguish if we are referring to the quality of being fair, to the concept of fairness, or to some Agent, involved in some Moral Value situation, playing the role of "fair" in a decision-making scenario (i.e. one of the many trolley dilemmas). Finally, the desire, also shared in this work, to modularise such a complex architecture into different ontological modules is unfortunately realised via the mere usage of different classes, subclasses of the same `owl:Thing` root node instead of creating modules and then importing them or reusing specific entities.

For the above-mentioned reasons, albeit the ambitious project and the prolific axiomatisation, the ontology model results in being a formalisation of fragments of several theories, but without further documentation of the resource it was unfortunately considered unsuitable for inclusion in the ValueNet ontology.

The next section presents the ValueNet ontology and the transposition of MFT and BHV theories as ontological modules, for which a more detailed theoretical description is provided in Sec. 3.2.4 and 3.2.2.

For a more in-depth analysis of the relationship between values and culture refer to Vauclair [334, 335]. For a history of values (at least from a western view perspective), refer to Edel [67]. For a broad social theory perspective on values, refer to Joas [158]. Finally, for an overview of the semantics of the term 'value' as a noun, or verb, and its elaboration in various theories of developmental psychology, see Rohan [272].

3.2 ValueNet Ontology

From the analysis of the state of the art, we can conclude that the value domain needs deeper investigation particularly from three perspectives:

1. values as foundations of a universal morality proper to human beings;

2. values as socio-behavioral norms/attitudes proper to every culture;

3. values as individual social expectations, realized in the daily behaviors of everyday life

ValueNet ontology is the module dedicated to the formalisation of moral values. Its structure and import and usage network are shown in Fig. 3.6. Values as foundations of a universal morality are investigated in MFT module, in Sec. 3.2.4. Values as cultural variables are investigated in BHV module, in Sec. 3.2.2. Values as individual behaviors and expectations are investigated in Sec. 3.2.6.

As stated in Chapter 1, the main assumption of this work is to treat modeled entities as conceptual frames, as intended in frame semantics. Therefore, being the ValueNet[4] modular ontology an extension of Framester hub, values are modeled as framal structures. It injects a constructionist view thanks to the reusage of the Description & Situation (DnS) [88] ontology design pattern, as described in Sec. **??**. Since each value is structured as a frame, it is declared triggered by other Framester entities, thus enabling a linguistic, cognitive, and factual grounding to values.

ValueNet purpose is twofold: (i) it aims at formalizing existing theories about moral and social values, to create a formal integrated environment, based on the general ValueCore module, described in section 3.2.1, which allows the integration of theoretical knowledge with experimental data based on a certain theory; (ii) it aims at operationalizing existing theories to develop sense-making tools, e.g., a value detector based on multiple theories, as explained in Section 3.3.

The first module presented in Sec. 3.2.1 is the ValueCore module. It models the minimum vocabulary to speak about the notion of *value*. Sec 3.2.2 and 3.2.4 describe in detail the Basic Human Values and the Moral Foundations Theory. The theoretical modules are followed by a description of their operationalisation, realised with the QUOKKA workflow as in Sec. 1.5. Sec. 3.2.6 focuses on "folk values", a less theoretically grounded module, but an attempt to inject even more commonsense knowledge in ValueNet ontology.

Finally, Sec. 3.2.7 describes morality as a compositional system, as stated by Curry [42].

[4] ValueNet repository is available here: `https://github.com/StenDoipanni/ValueNet`

The third and closing part of this chapter, Sec. 3.3, is dedicated to experiments on value detection from natural language.

We use the following prefixes to indicate:
`fschema:` = the Framester schema
`fs:` = Framestercore, the Framester core module
`vc:` = ValueCore, the ValueNet core module
`mft:` = Moral Foundations Theory module
`bhv:` = Basic Human Values module
`mm:` = Moral Molecules thoery module
`folk:` = Folk values module
`wn:` = WordNet resource
`vb:` = VerbNet resource
`db:` = DBpedia resource
`wiki:` = Wiktionary resource
`pb:` = PropBank resource
`fn:` = FrameNet resource
`dul:` = DOLCE Ultralight foundational ontology

3.2.1 ValueCore

The ValueCore module models the notion of *value* as a frame. It reuses the Constructive Description&Situation ontology design pattern [95][88], considering each value of each theory (formalized in separate modules.

The notion of *value* is a subclass of `fschema:ConceptualFrame`, in turn, subclass of `dul:Description`. It is satisfied by some `vc:ValueSituation`, namely, the realization/occurrence of some prototypical type of event involving some value. The class `dul:Description` is itself a subclass of `dul:SocialObject`, thus classifying the notion of value as a subclass of a "social object" in DOLCE. To provide an example, the *harm* MFT value, is modeled as `mft:Harm`, and it is an individual of the `vc:MFT_Value` class, subclass of the general `vc:Value` class. Being ValueCore the core module, it has the advantage to generalise the notion of value, and introduce a superclass covering all the specific notions of value, as intended in each theory, to cover every possible value situation.

From the literature analysis exposed in the previous section, the ValueCore module includes three main types of value-driven situations: (i) `vc:ValueAppraisal`: the appraisal of a situation performed by an agent, pivoted by a value; (ii) `vc:ValueCommitment`: the commitment of an agent to a value; and

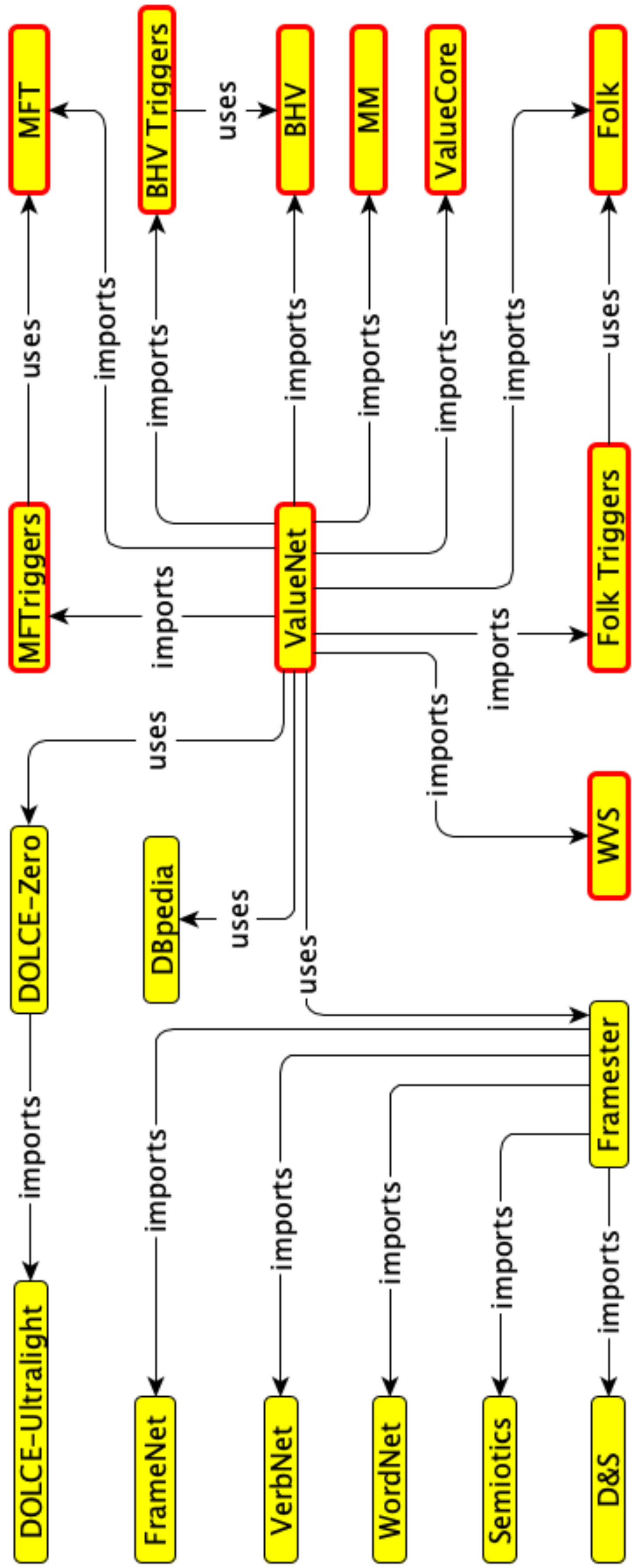


Figure 3.1: ValueNet import and usage network.

(iii) `vc:ValueRecognition`: the recognition, namely, the plain existence assertion, operated by some agent, of a value in some context. These three types of situations, modeled as framal structures, including necessary, optional, and external roles, allow modeling any type of event including some value, with increasing detail, proportional to the granularity of the scenario taken into consideration.
The ValueCore module can be explored online[5] or via the Framester endpoint[6].

ValueCore aims to answer the following competency questions:

- **CQ1**: What are the theories integrated into ValueCore? (namely, what are the subclasses of the `vc:Value` class?
- **CQ2**: What type of value situation is the one considered? Is it a ValueAppraisal? A ValueCommitment ?

CQ2 in particular is inferred automatically from text, as described in Sec. 3.3.

3.2.2 Basic Human Values (BHV) module

BHV and MFT as theories share some overlaps but start from quite different perspectives, greatly simplified: both theories propose a universal model, namely a model which should provide a cultural-agnostic explanation for the whole human value system, and for this reason they are modeled in ValueNet. But while MFT adopts a more developmental perspective (explained in detail in Section 3.2.4) and is more focused on morality, individuating dyadic universal foundations, BHV considers many socio-behavioral factors. This difference results in both theories having a relational opposition of values, but while MFT is organized in dyadic oppositions of one value and its violation, BHV circumplex model does not contemplate direct violations, but rather an opposition of behavioral focus and attitude.

The Theory of Basic Human Values (BHV) by Shalom Schwartz was proposed as a pan-cultural theory in the 1980s [288]. Its main assumption is that human values are organized in a value wheel, that is, an ordering that organizes values as a circumplex model, dividing them into four quadrants with two opposing axes, and a congruity continuum between adjacent values.

Schwartz and Bilsky [293, 294] suggested that all individuals must respond to three main human needs: (i) biology-based needs, (ii) social requirements for interpersonal interaction and coordination, and (iii) in-group social norms, as institutional demands, for

[5] The ValueCore module is available here:
https://raw.githubusercontent.com/StenDoipanni/ValueNet/main/ValueCore.ttl

[6] The Framester endpoint is available here: http://etna.istc.cnr.it/framester2/sparql

group welfare and survival. Values are therefore the cognitive representation of these needs.

The BHV framework is based on the following assumptions [291]:

- Human values are universal: All individuals, regardless of their culture or background, have the same basic values, which are innate and hardwired into the brain.
- Human values are organized hierarchically: The basic values form the foundation of all other values, and the relative importance of each value varies among individuals and cultures.
- Human values are related to individual and societal well-being: values prioritized by individuals and societies influence their overall well-being and quality of life.
- Human values are context-dependent: The relative importance of each value varies depending on the context, and individuals and cultures may prioritize different values in different situations.

Furthermore, cultural values reflect ideals that shape the beliefs and commitments of individuals and groups in the relation to the cultural environment [287]. They act as societal coping mechanisms to three situations of possible risks to a successful collaboration and co-location: (i) inside relations between individuals and groups; (ii) relations between individuals and the society as a cultural structure; (iii) individuals' interest vs the natural and social environment.

Originally, the model included 10 values [288], but, as shown in Fig. 3.2, the model was later refined to 19 values in total [296]. BHV relies on the opposition and similarity of values, grouped into macro-categories that are mostly determined by individual personality traits (self-transcendence vs self-enhancement, conservation vs openness to change). This model has inspired the design of a questionnaire (Portrait Values Questionnaire, PTV [303]) which has been employed by several studies to explore values across different countries [297]. In recent work [291], Schwartz provides evidence in favour of a pan-cultural arrangement of value priorities.

BHV has been tested on a vast number of subjects across 82 countries. However, one of the main criticism is its top-down approach, establishing the number and taxonomy of values a priori, and then validating it through dedicated experimentation.

In its 10 basic human values version, these are defined as follows, linguistic examples are taken from [297]:

- Power: the notion of recognised social status and prestige and, at the same time, power as control or dominance over people and resources (in the 2019 version these

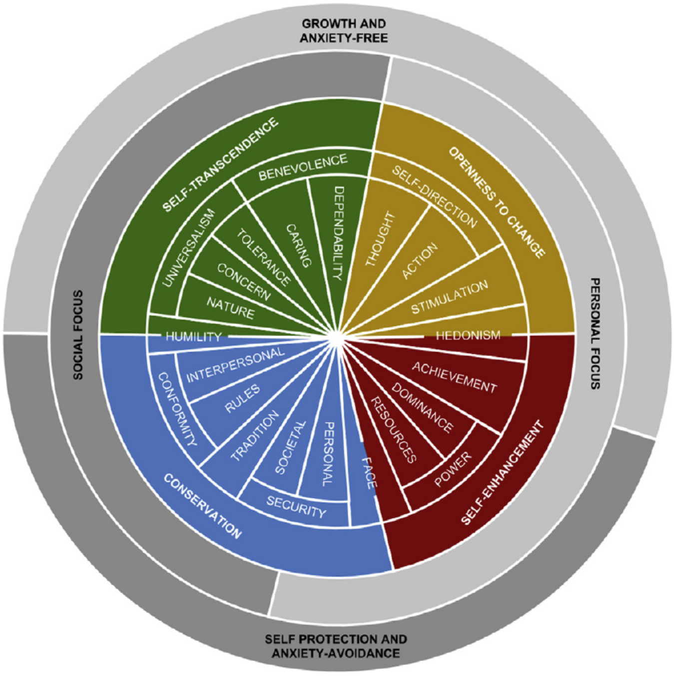


Figure 3.2: Basic Human Values circumplex model, image taken from [Giménez, August Corrons, and Lluís Garay Tamajón. *Analysis of the third-order structuring of Shalom Schwartz's theory of basic human values.* Heliyon 5.6 (2019): e01797.]

are two specifications of Power) e.g. 'He likes to be in charge and tell others what to do. He wants people to do what he says.'

- Achievement: personal success obtained via demonstrating competence according to certain social standards, e.g. 'Being very successful is important to him. He likes to stand out and to impress other people.'
- Hedonism: pleasure and sensuous gratification for oneself. e.g. 'He really wants to enjoy life. Having a good time is very important to him.'
- Stimulation: the need for excitement, novelty, and challenge in life, e.g. 'He looks for adventures and likes to take risks. He wants to have an exciting life.'
- Self-direction: freedom, independent thought and action, liberty to create, express and explore e.g. 'He thinks it's important to be interested in things. He is curious and tries to understand everything.'
- Universalism: understanding others, receiving appreciation, tolerance, and protection for the welfare of all people and nature. It underlies the notion of empathy, e.g. 'He thinks it is important that every person in the world should be treated equally. He wants justice for everybody, even for people he doesn't know.'

- Benevolence: protection, preservation, and enhancement of the welfare of people with whom one is in frequent personal contact, e.g. 'He always wants to help the people who are close to him. It's very important to him to care for the people he knows and likes.'
- Tradition: respect, commitment, and acceptance of the customs and ideas that traditional culture or religion provide the self, e.g. 'He thinks it is important to do things the way he learned from his family. He wants to follow their customs and traditions.'
- Conformity: restraint of actions, inclinations, and impulses likely to upset or harm others and violate social expectations or norms, e.g. 'He believes that people should do what they're told. He thinks people should follow rules at all times, even when no one is watching.'
- Security: being safe, harmonious, and stable of society, relationships, and self, e.g. 'The safety of his country is very important to him. He wants his country to be safe from its enemies.'

BHV Classes The ontology takes as source the BHV model reworked as in [108]. It is the attempt to formalize values as an *inner behavioral nudge*, related to outer stimuli, towards one (or more) of the four main axes as explained in the following.

The ontology includes 2 top classes representing the "attitude", i.e., a general view of the world, driving some more specific ordering principles; 2 classes representing a "focus", i.e., a taxonomical criterion that addresses the entities (social group, individual, society, class) supposed to profit the most from some value; 4 third-order clusters of values, which split the circumplex model into four quadrants, creating diagonal opposition and topical continuity; 12 second-order values, namely more specific clusters of values considering a more fine-grained granularity in framing events and situations of the world; and finally 19 first-order values, which explicitly state the patient/beneficiary of some value. We list here the ontological classes and axioms, from the most general ones (which in the circumplex model corresponds to most external sectors) to the most specific.
The highest-order layer of the circumplex model is formalized as follows:

- `bhv:GrowthAndAnxietyFree` : This is a pro-active attitude, characterizing a self-transcendent view of the world and a higher openness to novelty and change.
- `bhv:SelfProtectionAndAnxietyAvoidance` : this is a more reactive attitude, characterized by a self-centered view of the world fostering a closer and conservative attitude.

Note that, as shown in Fig. 3, the outer attitude ring and the focus one have no direct relation between them, being offset from each other, while the main four quadrants, and, as consequence, the single values, are instead axiomatized with restrictions on their attitude and focus. Moving therefore one ring inward into the circumplex model, the *focus* concept is specified in two classes and modeled as follows:

- `bhv:SocialFocus` : Focus on social issues and others than self, or focus on self, considered as a member of a social community. The focus expresses the main beneficiary of the behaviour determined by some Value e.g. the class `bhv:SelfTrascendence` is the superclass grouping all the Values having as focus society more than the individual;

- `bhv:PersonalFocus` : Focus on personal issues and self, both as a realization of self intended as freedom of thinking and action as well as dominance over others.

The third-order values layer structures the four main quadrants of the circumplex model. These are modeled as superclasses of more specific value situations, following Constructive Description and Situation pattern, considering more specific classes of situations as subclasses of more general ones, satisfying more specific descriptions, subclasses in turn of more general ones. Considering diagonal oppositions (meaning having an opposed value motivation), and according to their focus and attitude they are:

- `bhv:Conservation` : This macro category is focused on 'preserving stability and security', in particular 'with the emphasis on subservient self-repression, the preservation of traditional practices and protecting stability'. In the BHV ontological module `bhv:Conservation` class of value situations is axiomatised as:

```
SubClassOf :
((attitude some SelfProtectionAndAnxietyAvoidance) and
(attitude only SelfProtectionAndAnxietyAvoidance)) and
((focus some PersonalFocus) or (focus some SocialFocus))
```
(3.1)

Its opposite quadrant is:

- `bhv:OpennessToChange` : it consists in readiness for new experience, self centered values which foster physical and intellectual freedom and fulfillment. `bhv:OpennessToChange` class of value situations is axiomatised as:

```
                              SubClassOf :
((attitude some GrowthAndAnxietyFree) and
(attitude only GrowthAndAnxietyFree)) and
             ((focus some PersonalFocus) and
                (focus only PersonalFocus))
```
(3.2)

The sibling class to
`OpennessToChange` in the circumplex model is:

- `bhv:SelfEnhancement` : it consists in promoting self-interest, often at the expense of others, emphasising the search for personal success and dominance over others. `bhv:SelfEnhancement` class of value situations is axiomatised as:

```
                                              SubClassOf:
                   ((attitude some GrowthAndAnxietyFree) or
  (attitude some SelfProtectionAndAnxietyAvoidance)) and
(focus some PersonalFocus) and (focus only PersonalFocus)
```
(3.3)

In the opposed quadrant to `bhv:SelfEnhancement` there is:

- `bhv:SelfTrascendence` : it consists in promoting the well-being of society and nature above one's own interests, highlighting the acceptance of others as equals, as well as a concern for their well-being. `bhv:SelfTrascendence` class of value situations is axiomatised as:

```
                                         SubClassOf :
             ((attitude some GrowthAndAnxietyFree) and
             (attitude only GrowthAndAnxietyFree)) and
((focus some SocialFocus) and (focus only SocialFocus))
```
(3.4)

Finally, the full list of 19 first order BHV values is shown in Figure 3 and each value class is described in the OWL file[7].

[7] The ontology is available here: https://github.com/spice-h2020/SON/blob/main/SchwartzValues/ontology.owl

BHV Object Properties The object properties modeled in BHV module are:

- bhv:attitude : this property is used to declare the attitude corresponding to some values, namely bhv:SelfProtectionAndAnxietyAvoidance (re-active attitude) vs bhv:GrothAndAnxietyFree (pro-active attitude).
- bhv:focus : this property is used to declare the focus corresponding to some values, namely bhv:SocialFocus vs bhv:PersonalFocus.
- bhv:opposingFocus : serves the function of modelling oppositions, as described in the previous paragraphs and shown in Fig. 3.2.
- bhv:opposingValueMotivation : Following the polarity opposition Conservation vs OpennesToChange and SelfTrascendence vs SelfEnhancement, this property is used to axiomatise all the 4 third order classes of values declaring them as EquivalentTo : opposingValueMotivation some and opposingValueMotivation only the value in the opposite diagonal quadrant.
- bhv:panCulturallyMoreImportantThat : to express the eventuality of building a Pan Cultural Baseline For Values Priority.

BHV Competency Questions BHV module allows to answer some CQs according to BHV theory:

1. Is the entity *x* an instance of some value, according to BHV theory?
2. What values have as focus some bhv:SocialFocus or bhv:PersonalFocus?
3. What is the bhv:opposingFocus of some value?
4. What is the attitude of some value?
5. What is the opposing value motivation for some value?

3.2.3 BHV Triggers module

In this Section we present the operationalisation of BHV theory, namely the BHV Triggers graph. The graph is populated reusing semantic web resources' entities as triggers for Basic Human Values, and in particular here it is described the rationale determining the generation of the resource.

The BHV Triggers graph is realised using the QUOKKA workflow, as presented in Sec. 1.5. As the ImageSchemaNet graph (see Sec. 2.2.7) is a graph of semantic triggers for

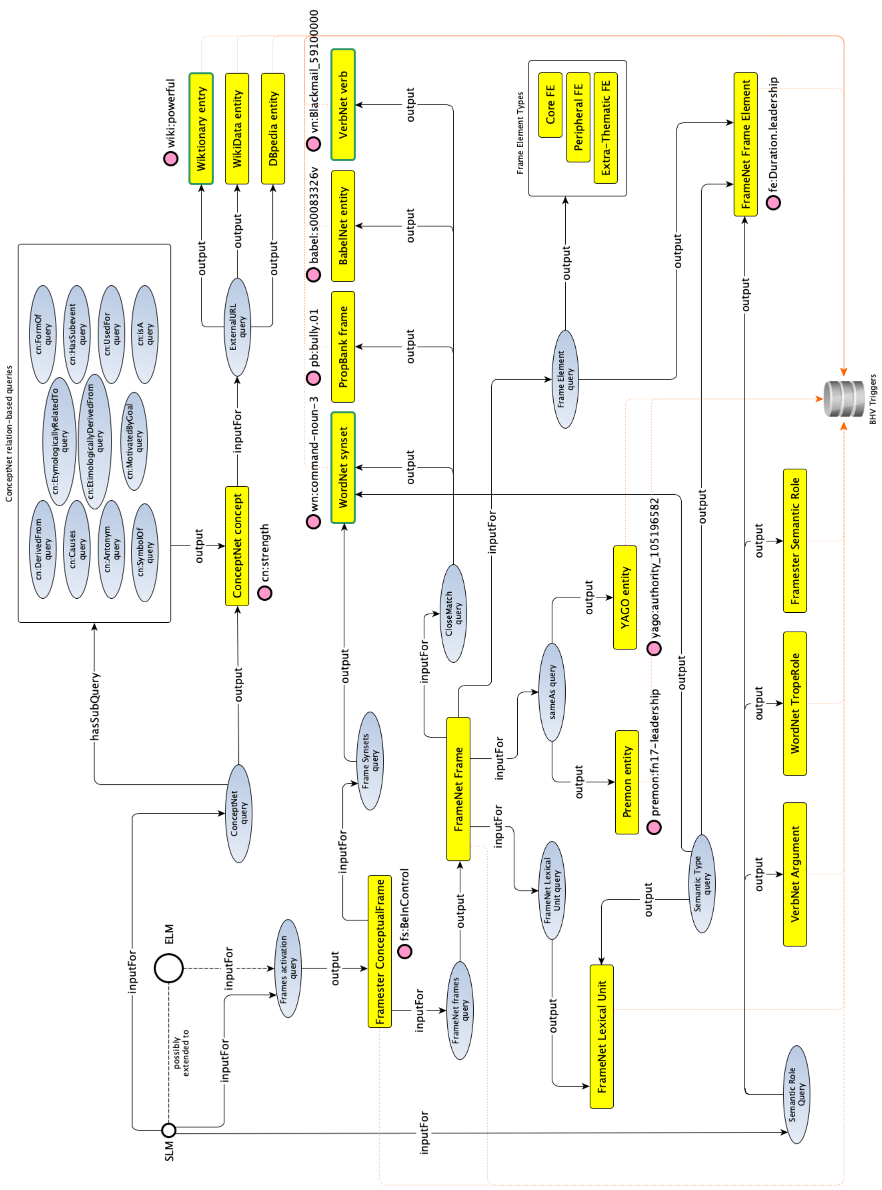


Figure 3.3: QUOKKA workflow applied to BHV. Individuals are taken from the bhv:Power value frame.

Words Selected to Represent Each Value: The Value Lexicon

Value	Lexical indicators for each value
Power	power, strength, control
Achievement	achievement, ambition, success
Hedonism	luxury, pleasure, delight
Stimulation	excitement, novelty, thrill
Self-direction	independence, freedom, liberty
Universalism	unity, justice, equality
Benevolence	kindness, charity, mercy
Tradition	tradition, custom, respect
Conformity	restraint, regard, consideration
Security	security, safety, protection

Figure 3.4: BHV lexicon used as Starting Lexical Material (SLM) for populating BHV triggers knowledge graph. Table taken from [12].

image schemas, BHV Triggers is a graph of semantic activators for the BHV value frame. It is the graph declaring the semantic triggering of entities introduced in the previously discussed theoretical modules.

The Starting Lexical Material (SLM) for BHV Triggers is provided by Bardi et al. [12], and it is available in Fig. 3.4. In Bardi et al., three terms closely related to the semantics of each value are selected.

According to this lexicalisation the terms are used as SLM in the QUOKKA workflow. As demonstrated, the lexicon is pretty small, and it is developed based on the 10 values version of BHV. BHV Triggers therefore, provides semantic triggers only for those 10, while BHV ontological module still models the 19 values version, as in Sec. 3.2.2.

The theoretical assumption is to consider each value - e.g. *Power* - as a conceptual frame. Consequently, as it happens in FrameNet, there is a set of lexical units evoking the `bhv:Power` frame, namely, the broad notion of *power*, according to the Basic Humans Value Theory. Fig. 3.3 shows, in pink dots, individuals of several classes triggering the `bhv:Power` value frame.

The number of semantic triggers per resource is available in the final Table 3.1.

The BHV Triggers knowledge graph population can therefore be summarised as follows:

- **Manual Lexical Units Selection**: For the `bhv:Power` value frame the SLM, as shown in Fig. 3.4, is constituted by "power", "strength" and "control". These elements are, therefore, given as input for the query expansion;

- **ConceptNet-driven Triggering**: ConceptNet allows the injection of commonsense knowledge in the resource. In the case of `bhv:Power` "power", "strength" and "control" are therefore passed to the next step, resulting, as shown in Tab. 3.1 in 130 concept entities retrieved as activators;

- **Wikidata and DBpedia triggering**: 142 entities are retrieved from the Wikidata dictionary, such as `wiki:powerful`, `wiki:empower`, `wiki:superpower`, etc. while 9 DBpedia entities are retrieved, such as `db:Biopower`, `db:Powerlifting` and `db:Horsepower`;

- **Frame-drive triggering**: the frames retrieved (and manually filtered) by the SLM for the `bhv:Power` value frame are: `fs:BeInControl`, `fs:Prohibiting`, and `fs:ManipulateIntoDoing`. Albeit several more frames were evoked these three are the only ones satisfying the theoretical criteria set by the BHV Power definition, namely referring in their semantics to situations subclasses of some `bhv:Power` situation. The presence of `fs:BeInControl` is quite straightforward, while, for the other two frames, the rationale was the following: (i) `fs:Prohibiting`: given a manual analysis of synsets subsumed by this frame, and its semantic roles, it seems plausible to say that, when it occurs a situation in which an entity defines a prohibition, there must be some power providing authority or strength to be able to "prohibit" something. For the (ii) `fs:ManipulateIntoDoing` the motivation was the same, to be able to manipulate something or someone to achieve a purpose, means to exert some (even if limited) power on the Undergoer entity;

- **Frame element-driven triggering**: Frame elements that the `bhv:Power` value frame inherits from the above-mentioned frames cover semantic roles such as `Degree.Power`, `Instrument.Power`, `Purpose.Power`, `Resulting_action.Power`, and many others.

- **FrameNet Lexical Units triggering** LUs from the above-mentioned FrameNet frames are declared as triggers for the `bhv:Power` value frame;

- **WordNet and VerbNet Lexical triggering**: due to the broad semantics of the above-mentioned frames, especially `BeInControl`, 1195 synsets from WordNet and 96 VerbNet verbs are declared as triggers, as shown in Table 3.1; the consistent amount of WordNet entities are motivated by their being retrieved also via the `skos:closeMatch` SPARQL query;

- **BabelNet triggering**: 931 BabelNet entities are retrieved as subsumed or having a `skos:closeMatch` to the above-mentioned frames;

- **Semantic Type triggering**: in contrast to Image Schematic entities, for which the semantic type SPARQL query is particularly productive, it is not productive for values. This is motivated by the ontological nature of the entities to be modeled, so it makes sense that for entities that are abstractions of sensorimotor experiences, a query expansion based on the ontological type is more productive than for entities whose very nature is not experienced through embodiment, but through social interaction.

We proceed now to describe the Moral Foundations Trigger and its ontological transposition, as well as the process and rationale for populating the knowledge graph operationalising it.

3.2.4 Moral Foundations Theory (MFT) module

The Moral Foundation Theory (MFT) [110] is proposed as a cultural-independent theory of moral and social values, inspired by Schweder's et al. work on universal human ethics [210]. For its social focus, it is tightly related to the investigation of moral emotions, with a particular focus on behavioural neuro-cognitivism. MFT is defined as 'a nativist, cultural-developmentalist, intuitionist, and pluralist approach to the study of morality' [110]:

- "nativist" in its neurophysiological grounding;
- "cultural-developmentalist" in including environmental variables in the morality-building process;
- "intuitionist" in declaring that there is no unique moral or non-moral trigger, but rather many patterns combining in a rationalized judgment;
- "pluralist" in considering that more than one narrative could fit the moral explanation process.

Differently from other value theories (e.g. Basic Human Values [296]) accused to be Atlanticist-centric and biased towards western values, its agnostic point of view towards cultural dependencies and individual values, is realized via its simple dyadic opposition structure. On one hand, the intension of value-violation dyadic oppositions is supposed to be culturally independent; on the other hand, their extension is dependent on the actual realisation of one (or more) dyadic value in some situation of the real world, meaning that e.g. the value of *Fairness* is per se universal, but what is considered to be fair is cultural, temporal and spatial dependent. The model proposed by [111] focuses mainly on single-value oppositions, where any pair of opposing values represent the poles of a prescribing/inhibiting dyad. At the core of MFT, there are six dyads of values and violations:

Table 3.1: BHV Values and amount of triggers from each semantic web resource involved in the value frame building process.

Value	Frame	Frame Element	WordNet	VerbNet	PropBank	ConceptNet	Wiktionary	WikiData	DBpedia	BabelNet	Umbel	YAGO	Premon
Achievement	0	14	147	7	3	40	29	0	0	129	1	106	1
Benevolence	2	34	244	36	7	26	26	1	1	128	2	66	3
Conformity	1	12	324	14	15	33	31	0	0	196	0	199	1
Hedonism	0	0	0	0	0	178	132	15	15	0	16	0	0
Power	3	36	1195	96	79	130	142	9	9	931	8	54	2
Security	2	27	279	17	12	127	133	13	13	212	23	180	1
SelfDirection	0	0	32	0	0	31	32	4	7	0	1	0	0
Stimulation	0	0	0	0	0	68	24	1	1	0	1	0	0
Tradition	1	6	50	0	0	46	41	2	2	37	1	29	1
Universalism	0	7	22	0	0	35	34	1	1	13	1	0	1

- *Care / Harm*: a caring versus harming behavior, it underlies virtues of gentleness, kindness, and nurture. It is grounded in attachment systems and some form of empathy, intended as the ability to not only understand but also feel, the same feelings as others, it allows us to imagine hypothetical scenarios, in which we are experiencing some positive or negative mental or physical state.

- *Fairness / Cheating*: this foundation is based on social cooperation and typical nonzero-sum game theoretical situations based on the evolutionary process of reciprocal altruism. It underlies ideas of justice, rights, and autonomy.

- *Loyalty / Betrayal*: this dyad is based on the positive outcome coming from a cohesive coalition, and the ostracism towards traitors. It is grounded in the clans and family-based dimension that for a long time characterized most of our tribal societies. The ability to create links and alliances was a way to increase the surviving percentage possibilities for oneself and his/her close group.

- *Authority / Subversion*: social interactions in terms of societal hierarchies, it underlies ideas of leadership and deference to authority, as well as respect for tradition. It is grounded in the hierarchical social interactions directly inherited by primates' societies.

- *Purity / Degradation*: it is grounded in the CAD triad emotions (Contempt, Anger, Disgust) and the psychology of disgust, it is one of the most spread dyadic oppositions, underlying religious (and not only) notions of living in an elevated, less carnal, more ascetic way. It underlies the idea of 'the body as a temple' which can be contaminated by immoral activities and it is foundational for the opposition between soul and flesh.

- *Liberty / Oppression*: it expresses the desire for freedom and the feeling of oppression when it is negated. It is grounded in feelings and experiences like solidarity, vs. episodes of unjustified violence or liberty restrictions.

Note that in 2022 version of the Moral Foundations Theory the positive pole `mft:Fairness` is splitted in two different subclasses: `mft:Proportionality` and `mft:Equality`.

Besides its relevance for the investigation of the emotional counterpart of value appraisal and the cross-cultural investigation of values, MFT has inspired the design of the Moral Foundation Dictionary [113] and, more recently, of the Extended Moral Foundations Dictionary [148], as described in Sec. 3.1.5, which combine theory-driven elements on moral intuitions with a data-oriented approach. Relationship with the emotion knowledge layer is described in Sec. **??**.

MFT Classes The MFT module is light-weighted considering the number of axioms, due to the fact that the whole theory is based on direct dyadic opposition of values and violations. MFT classes are:

- `mft:DyadicOpposition` : this is the superclass for all the value-violation dyads. It `dul:hasComponent` exactly 1 value and exactly 1 violation.
- `mft:Value` : this is the class for "positive" values shaping some behavior, it is subclass of `vc:Value` in the ValueCore module.
- `mft:Violation` : this class represent the violation to some value, and can also be conceived as "negative" value.

MFT Object Properties The object properties modeled in MFT module are:

- `mft:opposedTo` : some value is opposed to its violation in the dyadic structure. This property is symmetric.
- `mft:violates` : some violation violates some `dul:Norm`.
- `dul:hasComponent` : this property expresses the mereological aspect of some dyad.

MFT Competency Questions MFT module and (MFTriggers) allow to answer some CQs according to MFT theory, such as:

1. Is the entity x an instance of some value, according to MFT theory?
2. What is the value `mft:opposedTo` some entity x?
3. Is there some value in the sentence y?
4. What is the value profile of (namely the set of values activated by) some word or sentence?

3.2.5 MFTriggers module

In this section, we present MFTriggers, the new graph generated by reusing semantic web resources' entities as triggers for Moral Foundations Theory values, with a focus on how the generation of the resource is realized.

As for BHV in Sec. 3.2.3 each value (and violation) of each moral foundation, - e.g. *Harm*, *Authority*, etc. - is considered as a conceptual frame, consequently, as it happens in

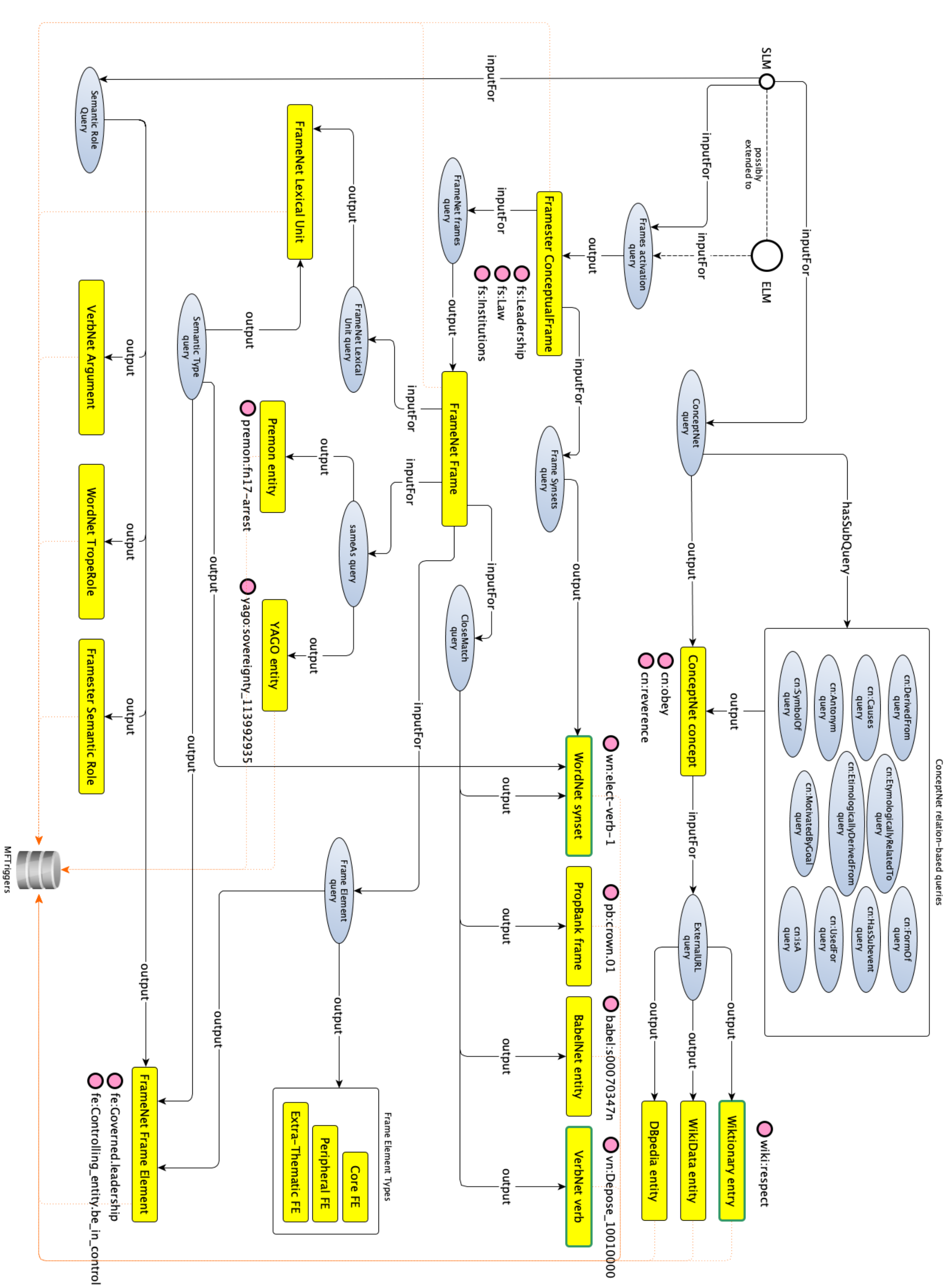


Figure 3.5: MFTriggers knowledge graph population reusing QUOKKA workflow. SLM material is provided by the Moral Foundation Dictionary [111].

FrameNet, there is a set of lexical units evoking the `mft:Harm` frame, namely, the broad notion of *harm*, according to the Moral Foundations Theory.

The MFTriggers resource is built using the QUOKKA workflow, describe in Sec. 1.5. The Starting Lexical Material (SLM) for MFTriggers module is provided by Frimer et al. [79] with the Moral Foundations Dictionary 2.0[8]. MFD2.0 is an extension of the original MFD [111], expanding the original lexicalisation for each pole of each MFT dyad. The coverage depends on the element, and it spans from a minimum of less than 50 lexical units for *Betrayal*, to more than 300 for *Harm* and *Authority*. On one side, therefore, the dimensions of MFT values frames depend on the amount of Starting Lexical Material (SLM) provided from the MFD2.0. At the same time, the dimension of the graph depends on the amount and dimension of the frames retrieved: semantically broader frames will subsume more synsets, verbs, and subframes, and be aligned with many entities from other resources.

The MFTriggers knowledge graph population is shown in Tab. 3.2. To provide insightful examples we focus here on the `mft:Fairness` value frame. Further details on the extraction and inclusion/exclusion choices are provided on the ValueNet repository[9]. MFTriggers can therefore be summarised as follows:

- **Manual Lexical Unit Selection**: As mentioned above, MFD2.0 is used as SLM for each MFTriggers value/violation entity. This allows diving more into the semantics of each MFT dyad. Some shades of meaning remain hidden from the mere value description. The notion of *Cheating*, for example, is lexicalised in MFD2.0 by some expected lexical units, such as "betray" and "inequities", but also by some other terms, polarised to more specific nuances such as "bigot" or "racism". This points out once again the matter of lexical units as "carving out some portion of reality". The Fairness/Cheating dyad, in fact, is unbalanced in its semantic load: actually, when referring to this particular dyad, a more precise meaning would be caught by the label `Fairness / ¬Fairness`. The lexicalisation of MFD2.0, which refers to this dyad as "Fairness.virtue / Fairness.vice", is biased towards the fairness pole, and uses the "cheating" label improperly, since it only overlaps partially with the set of situations violating or negating fairness. It makes sense that a situation in which it is spotted, expressed, appraised, or identified as some form of racism, can be morally loaded with the negation of fairness.

- **ConceptNet-driven triggering**: the amount of triggers retrieved from ConceptNet is, by design of the QUOKKA workflow, directly depending on the number of units

[8] The Moral Foundations Dictionary 2.0 is available here: https://osf.io/ezn37/

[9] The MFTriggers and BHV Triggers modules are available and documented here: https://github.com/StenDoipanni/ValueNet

in the SLM, Tab. 3.2 shows numbers in detail;

- **Wikidata and DBpedia triggering**: keeping the Fairness case, 188 Wiktionary entries are retrieved, such as wiki:equality, wiki:justice, and impartial. As for the factual knowledge triggers, there are 26 entities from DBpedia, such as db:Honesty, db:Rights, and db:Sportsmanship;

- **Frame-driven triggering**: the results for this query vary greatly depending on the value in question: the peak is reached with the mft:Harm value frame (21), while in the cases of mft:Liberty and mft:Betrayal no frame appropriately covering the semantics of these values is retrieved. This might stimulate reflections on the nature of the formalization of resources such as FrameNet, which seem particularly florid with frames related to the notion of *harm*. To keep the mft:Fairness example, some of the frames retrieved as fairness situations are fs:Legality, fs:Trust, and interestingly fs:Revenge. It seems, in fact, from MFD2.0 , which includes the "revenge" lexical unit in SLM, that conceptually the notion of *revenge* implies some twisted notion of fairness, it underlies the idea of 'eye for an eye".

- **Frame element-driven triggering**: the mft:Fairness value frame inherits semantic roles such as fe:Degree.Fairness, fe:Espressor.Fairness, fe:Judges.Fairness, Source_of_authority.Fairness, and many others;

- **WordNet and VerbNet lexical unit triggering**: 494 WordNet synsets and 27 VerbNet verbs are declared triggers of the mft:Fairness frame;

- **BabelNet triggering**: thanks to the skos:closeMatch query 322 triggers are related to the mft:Fairness value frame;

- **Premon triggers**: the Premon entities are directly dependent on the amount of FrameNet frame retrieved and declared as triggers, in this case, 9.

The next section is focused on a bottom-up approach: by investigating a cultural relativist position about those that we have called Folk Values.

3.2.6 Folk Values module

Haidt and Hersh [128] interviewed liberals and conservatives in a study about homosexuality in the US, and found that the affective reactions to some stories were more accurate predictors of their moral stance than their claims about positive/negative consequences. Haidt's social intuitionist model injects the moral dimension and moral judgments into the process of action evaluation. Haidt [125] asserts that in every society people make

Table 3.2: MFT Values and amount of triggers from each semantic web resource involved in the value frame building process.

Value	Frame	Frame Element	WordNet	VerbNet	PropBank	ConceptNet	Wiktionary	WikiData	DBpedia	BabelNet	Umbel	YAGO	Premon
Fairness	11	93	493	27	16	116	188	10	26	322	11	177	11
Cheating	9	90	458	50	38	236	390	18	28	334	17	281	9
Liberty	0	0	32	0	0	116	129	8	11	0	3	0	0
Oppression	3	29	111	13	8	69	64	3	8	56	1	32	3
Care	6	58	795	101	22	1622	1421	43	45	360	43	255	5
Harm	21	330	2211	501	268	288	535	24	61	1605	18	837	28
Loyalty	4	37	416	28	10	143	287	24	27	264	17	128	4
Betrayal	0	0	27	0	0	49	84	4	5	0	3	0	0
Authority	12	107	4978	97	91	271	520	31	38	3633	22	905	11
Subversion	4	41	371	39	16	130	238	11	23	229	5	178	4
Sanctity	7	52	498	22	20	272	579	53	85	417	33	240	7
Degradation	7	73	189	46	24	388	764	38	42	124	30	47	7

evaluations about anyone or any matter respecting (or not) a "virtuous behavior" with respect to a certain social category.

These social-standard-determined virtues are mandatory in the sense that everyone is expected to respect them, and is expected to be judged if not. People who fail or refuse to do so are subject to social criticism or punishment. Vauclair [334] defines moral judgments as 'evaluations (good vs. bad) of the actions or character of a person that are made for a set of virtues held obligatory by a culture or subculture'.

The existence of this module is motivated therefore by a factual and pragmatic approach.

It is, in fact, true that, albeit the huge debate about *what* and *how many* moral values are there, people still have commonsense knowledge about behaviors that shape everyday social interaction.

For instance, adopting or not a healthier, low-fat diet may not be considered directly a moral issue such as trolley dilemmas, but many subcultures focused on the `mft:Sanctity` pole of the MFT dyad, resonating with the 'my body is a temple' metaphor, meaning it is morally inferior not to adopt healthy habits [312]. In everyday life, people can answer questions like 'what is that you look for in a good friend?' or 'what do you evaluate the most in your search for a soulmate', etc. Ultimately, BHV module aims at covering individual and cultural values, MFT is focused on the moral judgment aspect of values, and the Folk module deals with this aspect of social expectation.

Therefore, to investigate culturally relative morality, we tried to reverse engineer this big substratum of commonsense knowledge.

The generation of the Folk value module, and its operationalisation with a bottom-up approach, is done by following these three steps:

1. Scrape the web for extended lists of what *people* consider values. Note that this could be any kind of list, from online life coach guidelines to live a better life, to unofficial repositories of cultural values, that we named Folk Values;

2. Model them in a dedicated ontological module;

3. Establish a taxonomy among them, and filter the granularity of detail, namely, remove duplicates with different names (e.g. "Victory" and "Winning") which were pointing at a very similar portion of reality.

4. Treat those values as frames, therefore as classes of situations for which it is possible to individuate roles, lexical triggers, and factual entities that, in their semantics, point at a folk value-related occurrence of a certain situation.

While this ontological module does not bear any domain expert authority, its intention is exactly to provide, next to the theoretical ontological transposition, a module that considers also a bottom-up determined folk perspective, and which allows spotting more cultural depending entities. Among gathered folk values, relevant retrievals not considered in BHV and MFT modules, and related to a much more pragmatic dimension, are e.g. `folk:fitness`: the social importance of being fit; `folk:punctuality`: social appraisal related to not being late; `folk:frugality`: moral judgments about bragging about wealthiness; `folk:wealthiness` itself; `folk:authenticity`: the idea of being sincere in everyday manifestation and "not interpreting a character"; `folk:Intelligence`: being above average in commonsense intelligence-related tasks; and many others, for a total amount of more than 200 folk values, formalised as frames.

The knowledge graph population process is realised via applying again the QUOKKA workflow as in Sec. 3.2.3 and 3.2.5. The full table with Folk values triggers divided per resource provenance is available on the ValueNet GitHub[10] and it is integrated into the Framester ontology, and queryable from the Framester endpoint.

3.2.7 Moral Molecules module

The Moral Molecules (MM) module is the ontological transposition of Curry's work [42] focused on Morality as Cooperation. In this view morality is said to be: 'the label that philosophers and others have attached to these cooperative solutions'. Therefore moral elements are derived by adherence to 7 cooperation strategies, while negative moral elements are constituted by the violation of the above-mentioned strategies.

Here we list the moral elements, the corresponding social cooperation strategy grounding it, the social problem to which it is declared a strategy for, the underlying virtues, their transgressions as vices, and even an epithet provided in [42] and [43], to ground moral elements in commonsense knowledge.

- *Family Values*: it is grounded in the `mm:Kinship` social cooperation strategy, its basic idea is to give family special treatment. It underlies the idea of having a duty to care, and special obligations towards kin. The social situation strategy is "kin altruism", while the violations of this element are said to be neglecting care for kin, or occurrences of incest. The epithet provided to exemplify this moral element is 'Blood is thicker than water';

- *Group Loyalty*: grounded in Mutualism cooperation strategy, its basic idea is to work together rather than work alone. Its virtues are loyalty, unity, and solidarity,

[10]The Folk module in the ValueNet GitHub is available here: https://github.com/StenDoipanni/ValueNet/tree/main/ThatsAllFolks

but also conformity to in-group social rules. Social opportunity is any occurrence of Coordination, and its transgressions are betrayal and treason. The epithet provided is 'United we stand, divided we fall';

- *Reciprocity*: it is rooted in the Exchange strategy, whose basic idea is to return both favours and injuries. Its virtues are reciprocity, trustworthiness, and forgiveness, while its vices are cheating and ingratitude, realised in social dilemmas scenarios. The epithet is 'One good turn deserves another';
- *Heroism*: it is grounded in a Hawkish attitude, namely publicly demonstrating power. The underlying virtues are bravery, fortitude, and largesse, while vices are cowardice and miserliness, displayed in conflict resolution scenarios. Its epithet is 'With great power comes great responsibility'.
- *Deference*: opposite to the previous one, its cooperation strategy is based on demonstrating a Dovish attitude, namely being submissive to superiors. Related values are respect, humility, awe, and reverence, while their opposite vices are disrespect and hubris. The situation for its occurrence is the same conflict resolution scenario as in Heroism, while its epithet is 'Blessed are the meek';
- *Fairness*: it is grounded in the value of equal distribution of disputed resources and it is violated by unfairness and favouritism. The social situations for its occurrences are conflicts whose nature reside in bargaining matters, and the epithet is 'Let's meet in the middle';
- *Property Rights*: its basic idea is the recognition of possession of goods or resources, and in particular respect first possession. Its virtues are: respect for property and property rights, while, consequently, its violations are theft and trespassness attitudes. It occurs in conflict resolution scenarios whose nature resides in ownership recognition. Its epithet is 'Possession is nine-tenths of the law'.

MM classes The MM module includes the following main classes:

- `mm:MoralElement`: The class representing "atoms of morality" as explained above. It is axiomatised as:

```
EquivalentTo:
(basedOnCooperationStrategy some CooperationStrategy) ,
(hasNegativeCounterpart some NegativeMoralElement)
```
(3.5)

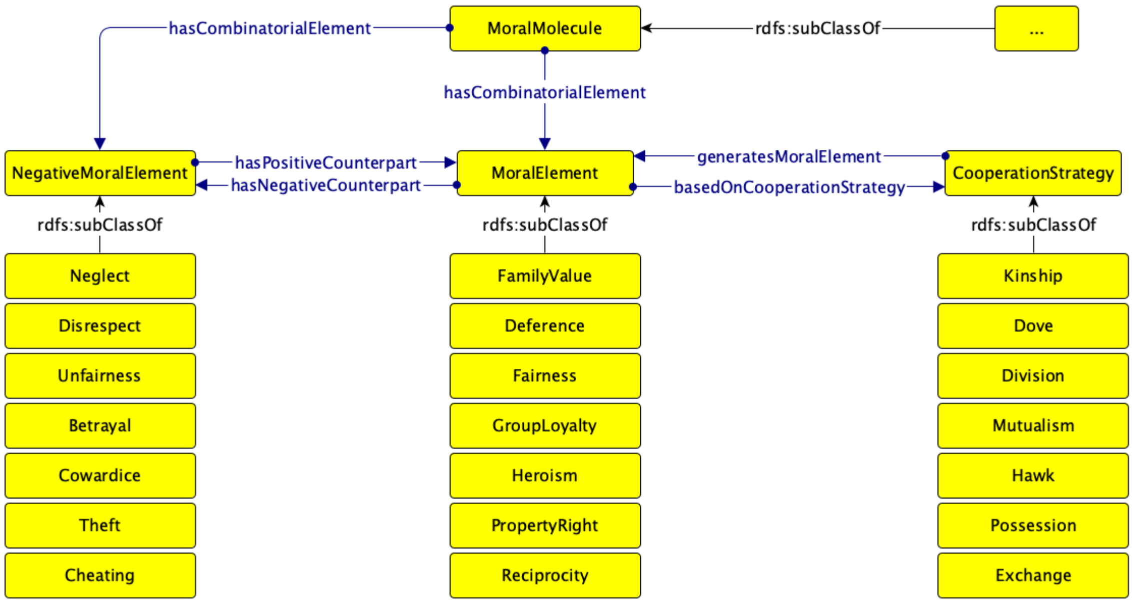


Figure 3.6: Moral Molecules basic structure.

It is the superclass of: `mm:Deference`, `mm:Fairness`, `mm:FamilyValue`, `mm:GroupLoyalty`, `mm:Heroism`, `mm:PropertyRight` and `mm:Reciprocity`. Each of these moral elements is in turn axiomatised as having its specific counterpart, and being based on a specific cooperation strategy. For example `mm:Fairness` is axiomatised as:

```
EquivalentTo:
(basedOnCooperationStrategy some Division) ,
(hasNegativeCounterpart some Unfairness)
```
(3.6)

- `mm:NegativeMoralElement`: the class to represent vices and violation of a certain `mm:MoralElement`. It is axiomatised as:

```
EquivalentTo:
(hasPositiveCounterpart some MoralElement)
```
(3.7)

It is the superclass of `mm:Betrayal`, `mm:Cheating`, `mm:Cowardice`, `mm:Disrespect`, `mm:Neglect`, `mm:Theft` and `mm:Unfairness`. Each negative moral element in turn is axiomatised declaring its specific positive counterpart, for example `mm:Unfairness`:

```
EquivalentTo:
(hasPositiveCounterpart some Fairness)
```
(3.8)

- mm:CooperationStrategy: the class to represent the 7 types of cooperation grounding the moral elements. It is axiomatised as:

$$\begin{aligned}&\texttt{EquivalentTo:}\\&\texttt{(generatesMoralElement some MoralElement)}\end{aligned} \tag{3.9}$$

  It is the superclass of mm:Division, mm:Dove, mm:Exchange, mm:Hawk, mm:Kinship, mm:Mutualism and mm:Possesion. Each cooperation strategy is in turn axiomatised for its specific moral element, e.g. for mm:Division:

$$\begin{aligned}&\texttt{EquivalentTo:}\\&\texttt{(generatesMoralElement some Fairness)}\end{aligned} \tag{3.10}$$

- mm:MoralMolecule: this class represents the moral profile of a moral complex situation, obtained via combining moral elements in more complex aggregates. It is axiomatised as follows:

$$\begin{aligned}&\texttt{EquivalentTo:}\\&\texttt{(hasCombinatorialElement min 2 (MoralElement or NegativeMoralElement))}\end{aligned} \tag{3.11}$$

  We automatically generated the possible unique combinations and declared them as subclasses, for a total of 16.637 classes. Then we declared the examples mentioned in [42] as individuals of the class combining the elements composing the complex situation. For example mm:Fraternity is a mm:FamilyValue_GroupLoyalty. Note that, according to MM theory, this formalisation does not explicit what kind of relations are acting among moral atoms, therefore per absurdum, we could have Fraternity, roughly conceptualised as 'behaving in the same way with someone non kin as it was kin', axiomatised in FOL as:

$$\begin{aligned}&\texttt{kinAttitude(X,Z)} \wedge \texttt{kinTo(X,Z)} \wedge \texttt{sameGroup(X,Y)} \wedge\\&\texttt{kinAttitude(X,Y)} \wedge \neg\texttt{kinTo(X,Y)} \Longrightarrow\\&\texttt{fraternityAttitude(X,Y)}\end{aligned} \tag{3.12}$$

  and some quite different concept such as "ClanConflict", roughly defined as 'A guides clan X and is kin to B, who guides clan Y; A and B fight, therefore also X and Y fight':

$$\texttt{clanLeader(A,X)} \wedge \texttt{clanLeader(B,Y)} \wedge \texttt{fight(A,B)} \implies \texttt{fight(X,Y)} \tag{3.13}$$

And both of them would probably be occurrences of situations including the moral atoms `mm:GroupLoyalty` and `mm:FamilyValue`.

MM Object Properties The main properties are those already mentioned in the previous paragraph axioms:

- `mm:basedOnCooperationStrategy`: it takes as domain some `mm:MoralElement` and some range some `mm:CooperationStrategy`;
- `mm:generatesMoralElement`: the inverse of the previous mentioned property, it takes as domain some `mm:CooperationStrategy` and as range some `mm:MoralElement`;
- `mm:hasCombinatorialElement`: the property used to declare the moral profile of some situation; it takes as domain some `mm:MoralMolecule` and as range some `mm:MoralElement` or `mm:NegativeMoralElement`;
- `mm:hasNegativeCounterpart` and `mm:hasPositiveCounterpart`: properties used to relate specific moral elements individuating a dyad virtue-vice;
- `mm:violatesCooperationStrategy`: the property used to close the triangle of relations among coop strategies, virtues and vices, it takes as domain some `mm:NegativeMoralElement` and as range some `mm:CooperationStrategy`.

Note that, there are some partial overlaps among MM theory, BHV, and MFT (we do not consider Folk values since these values are scraped bottom-up and not organized top-down in a theory). The main problem is that being declared as strategies to resolve social conflicts or shaping social interactions, spotting them from natural language would require being able to detect patterns of interactions that, at the actual state of the art, with existing frames, is hard to realise. Further development, MM operationalization, and moral profile extraction adopting the combinatorial structure provided by this theory, are envisioned in this Chapter Conclusions.

MM Competency Questions MM module allows to answer some CQs according to MM theory, such as:

1. What is the negative counterpart of some moral element?
2. What cooperation strategy is based on the e.g. Deference moral element?
3. What are the atoms (moral elements) for some morale molecule?

The MM module is available on the ValueNet GitHub[11].

The next section describes experiments of automatic value situation detection from text, reusing well-known resources and showing how the operationalisation of the ontological module allows a completely explainable pipeline of extraction of value from text. This method allows keeping a trace of the *triggering locus*, namely the element of the sentence triggering some value, and at the same time, it exploits organised semantic information, being able to extract further knowledge from semantic dependencies expressed in natural language.

[11]The Moral Molecules module is available here: `https://github.com/StenDoipanni/ValueNet/tree/main/MoralMolecules`

3.3 ValueNet Evaluation

This section describe some experiments done to test and evaluate the ValueNet ontology. It is organised as follows: Sec. 3.3.1 shows how it is possible to exploit BHV and MFT theories, used together, thanks to their ontological modules in ValueNet, to infer knowledge about individual values and their being in conflict with, or shared by, some other individual. The use case scenario is taken from the Modern Art Gallery (GAM) in Turin, and it comes from the H2020 SPICE project linked data hub. Sec. 3.3.2 instead provides two experiments of value detection from natural language, the first one is developed using the Moral Foundations Twitter Corpus and it is tested only with regard to MFT values detection. The second experiment is done on the Moral Foundations Reddit Corpus and it shows data from MFT, BHV and Folk values extraction.

3.3.1 BHV and MFT Inference Testing

In this first experiment we provide test the inference power of the BHV and MFT ontological module. This experiment has been published in [51], with best paper nomination.

BHV and MFT describe primitive framing of values as descriptions, and are typically associable to real world occurrences (situations), named, in the ValueCore module, as `vc:ValueSituation`. A value situation presents elements coherent to the conceptualization of BHV or MFT, so that it can answer competency questions mentioned in Sections 3.2.2 and 3.2.4.

To show some possible inferences exploiting MFT and BHV axiomatisation, we propose here a scenario answering CQs mentioned in Section 3.2.2 and 3.2.4, involving at the same time three types of value situations according to ValueCore module, namely `vc:ValueRecognition`, `vc:ValueAppraisal` and `vc:ValueCommitment`.

Value Scenario UserA and UserB are visiting an art gallery and see a painting depicting Pietro Micca ("Pietro Micca nel punto di dare fuoco alla mina volge a Dio e alla Patria I suoi ultimi pensieri" - "Pietro Micca, the moment before setting fire to the bomb, directs his thoughts to God and his motherland") by Andrea Gastaldi. Pietro Micca is described as an Italian patriot who gave his life to save the to-be-born state of Italy, igniting some dynamite to detonate a tunnel that was being invaded by enemy soldiers.

Pietro Micca's action can be modeled as a `vc:ValueCommitment` situation, nested in two different interpretations of UserA and UserB which can be modeled as `vc:ValueRecognition` situations, and for each of them would be possible to express the appraisal and the desirability of some action for both Users in a `vc:ValueAppraisal` situation[12].

BHV Inferences UserA declares to be proud of the action made by Pietro Micca, sharing with him the value *Patriotism*. UserB disagrees considering more important *Self Preservation* than sacrificing one's own life to defend the country. Thanks to BHV module and the lexical tokens linked to the first order values, *Patriotism* is inferred as being an instance of both `bhv:Societal` and `bhv:Caring` (see Sec. 3.2.2 CQ1), subclass of `bhv:Security` and `bhv:Benevolence`. Therefore, having as opposing value motivations (namely being in the quadrant opposed to) both `bhv:SelfEnhancement` and `bhv:OpennessToChange` (see Sec. 3.2.2 CQ5). In contrast *Self-Preservation* is an instance of `bhv:Action`, subclass of `bhv:SelfDirection`.
We can infer that UserA's instance of *Patriotism* has `bhv:focus` some `bhv:SocialFocus` (see Sec. 3.2.2 CQ2) and attitude both `bhv:SelfProtectionAndAnxietyAvoidance` and `bhv:GrowthAndAnxietyFree` (see Sec. 3.2.2 CQ4); while for UserB's value instance we can infer that it has some `bhv:PersonalFocus`, opposed to UserA's focus (see Sec. 3.2.2 CQ3) and `bhv:GrowthAndAnxietyFree` attitude.

Similar scenarios to the one proposed here in natural language are available serialized in turtle syntax both on the ValueNet and SPICE project GitHub[13].

MFT Inferences UserA declares to be proud of the Action made by Pietro Micca, focusing on the result of this action, namely the Liberty of Italy. UserB disagrees, considering more important Pietro Micca's life than any victory in war, in fact she/he considers it useless to sacrifice oneself for any country. Thanks to MFTriggers 'LibertyOfItaly' is inferred as triggering a `mft:Liberty` value Situation and 'PietroMiccaSacrifice' is inferred as triggering an `mft:Harm` situation (see Sec. 3.2.4 CQ3-CQ4). Thanks to the MFT dyadic model, 'LibertyOfItaly' is inferred as being an instance of `mft:Liberty` (see Sec. 3.2.4 CQ1), while 'CareOfPietroMicca' is an instance of `mft:Care`, being opposed to 'PietroMiccaSacrifice', which is an instance of `mft:Harm` (see Sec. 3.2.4 CQ2).

[12]We do not provide details about the ValueCore possible inferences here since it's not the main focus, but further details are available on the ValueNet GitHub:
`https://github.com/StenDoipanni/ValueNet`
and on the SPICE project GitHub:
`https://github.com/spice-h2020/SON/blob/main/SchwartzValues/Schwartz_scenario.ttl`
[13]The SPICE project GitHub is available here: `https://github.com/spice-h2020/SON`

3.3.2 Frame-based Moral Value Detection

In this second section dedicated to ValueNet testing we focus on the extraction of value knowledge from natural language.

In our work we aim at addressing critical issues, such as automatic moral value detection, in the most versatile and transparent way and, to the best of our knowledge, the approach we propose is unprecedented in moral values detection.

Most of value detectors are based on neural models and provide a single tag for each sentence, which can be particularly problematic with more complex sentences that include different aspects of the same or multi-faceted matter. It is in fact the case of real world commonsense knowledge: even considering a small exchange of a few tweets on the web, many different values can be taken into the discourse, and a single tag for the whole content is often not sufficient nor satisfying of the latent semantic complexity.

Our Value Detector tool - whose beta version is available online[14] - is composed of the following three main steps, as described in Sec. 1.6.1: (i) it takes as input a sentence in natural language, this sentence is passed to FRED tool. FRED parses the sentence and, considering its syntax, builds a knowledge graph of semantic dependencies, performing frame extraction, WordNet and VerbNet disambiguation, entity recognition etc.

The second step (ii) consists in navigating on the graph produced by FRED and, out of all the subjects, predicates, and objects of all triples, namely all nodes and arches in the graph, take into account only those referring to entities retrieved from reused resources - namely FrameNet, WordNet, VerbNet, DBpedia entities etc. leaving apart all the others. For each of these nodes the Value Detector performs a SPARQL query to ValueNet resource, to check if it is present any semantic trigger of some value.

Finally, (iii) for each triggering occurrence successfully retrieved, a triple is added to the original graph declaring the triggering. A final "valueprofile.ttl" file is produced as output, containing the graph generated automatically by FRED and all the value activation occurrences localised on the graph.

Note that factual situations can evoke some Value, and be opposed by some Violation, creating multi-shaped scenarios, in which the same sentence describing a complex event or action or entity can evoke different Values and their Violations at the same time.

In order to show the capabilities of the frame-based Value Detector let's consider two non-trivial plausible examples: (i)'Police brutality is a threat to people's trust in institutions.' and (ii)'If governments act against democracy we must protest and even fight to defend freedom.' taken from online news recent debates.

As shown in Fig. 3.7 several lexical units are aligned / disambiguated to semantic web

[14] The Value Detector is available in its online beta version here http://framester.istc.cnr.it/semanticdetection/values

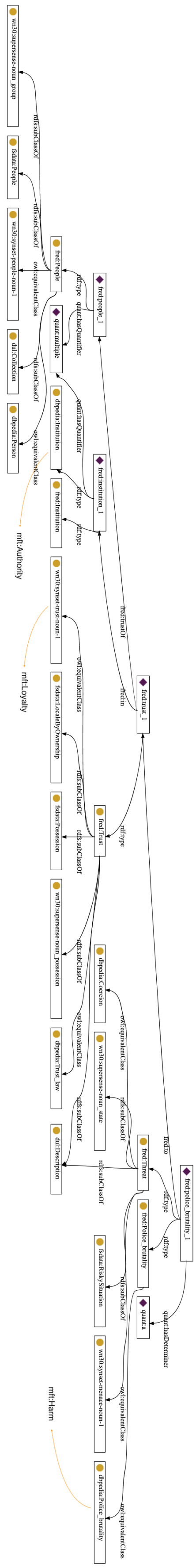


Figure 3.7: Value Detection from FRED graph produced for the sentence: 'Police brutality is a threat to people's trust in institutions.'

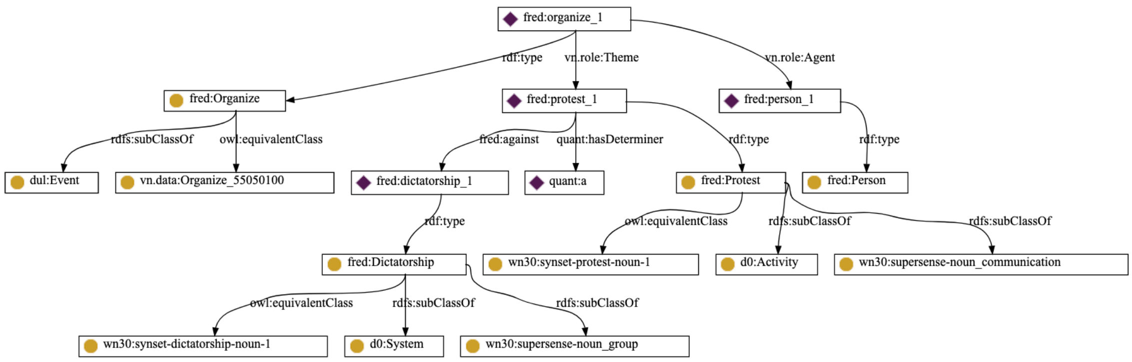


Figure 3.8: Value Detection from FRED graph produced for the sentence: 'If governments act against democracy we must protest and even fight to defend freedom.'

resources. Those retrieved as value triggers here are: `db:Institution` triggering `mft:Authority`; `db:Police_brutality` triggering `mft:Harm`; and `wn:synset-trust-noun-1` triggering `mft:Loyalty`. Therefore the value profile of this sentence includes three occurrences of two values and one violation (namely two positive and one negative value). Note that this frame-based value detection is explainable, in the sense that it allows to backtrack the localisation of some value referring to the precise lexical unit in the sentence, in contrast to blackboxes neural models. Furthermore it also allows to navigate on the graph maintaining semantic trace of the dependencies of value triggers. Fig. 3.7 in fact shows that the `mft:Harm` trigger ("Police brutality") is directed towards something ("trust") that triggers the positive value `mft:Loyalty`, which in turn is loyalty towards the `mft:Authority` trigger.

Considering the other example (ii) 'If governments act against democracy we must protest and even fight to defend freedom.' the profile is far more complex. Fig. 3.8 shows the semantic dependencies, while the value profile of this sentence includes: entities from DBpedia such as `db:Democracy` triggering `mft:Fairness` and `db:Government` triggering `mft:Authority`; frames: `fs:Defend` triggering `mft:Care` and `fs:HostileEncounter` triggering `mft:Harm`; VerbNet entries: `vn:Defend_85000000` triggering `mft:Care` and `vn:Protest_71000000` triggering `mft:Subversion`; and finally WordNet synsets: `wn:contend-verb-6` and `wn:fight-verb-2` triggering `mft:Harm`, `wn:freedom-noun-1` triggering `mft:Liberty`, and `wn:protest-verb-1` triggering `mft:Subversion`. Coherently to the more intense value detection also the complexity of the semantic structure of the graph increases: some action, having as VerbNet role Agent the trigger of `mft:Authority` ("Government") is against an entity trigger of `mft:Fairness`. This same action entails something triggering `mft:Subversion` ("protest"), performed by some Agent which is also the actor of two other actions: one ("fighting") that triggers `mft:Harm` and another ("defend") that triggers `mft:Care`.

The analysis conducted here can be verified giving as input the above mentioned sentences to the online demo of the value detector.

It is necessary to clarify, first of all, a point that still seems unaddressed by the main works proposing annotated corpora with values: the meaning of annotation. Although important sources such as the Moral Foundations Twitter Corpus (MFTC), used in the first experiment, and the Moral Foundations Reddit Corpus (MFRC) used in the second experiment, provide extensive explanations of the origin and composition of the corpus, and devote attention to the foundational theory for annotation (MFT) and its nuances, it remains obscure what principle is required to perform the annotation task. Are annotators required to identify all values? to identify those for which there is an explicit lexical trigger? are annotators required to indicate only the most *salient*? This question is of no minor importance, for as we shall see, the tool proposed here is not merely a classifier, but rather a value situation analyzer. This means that if the annotator is asked to indicate the prominent value of a certain sentence, it will bypass minor, but still essential, value elements for understanding the sentence. This is what the system proposed by this paper does: identify all those elements whose semantic load is connoted with value knowledge, and then exploit the graph representation of the semantic dependencies of the sentence for further inference. To provide a simple example, the phrase 'The disgusting betrayal will be punished' could be tagged as `mft:Betrayal`, since the lexical trigger explicitly refers to an occurrence of betrayal. But it could also be tagged with `mft:Loyalty`, since, applying a cognitive process based on common-sense knowledge, if something is defined as *disloyalty*, and it is configured as an action deserving of punishment, a negative epistemic stance towards the occurrence of betrayal can be derived from this, and thus a commitment to what in MFT theory is its opposite. The experiments described in the following sections are devoted to the detection of values from text, showing which and how many values are bypassed by a single sentence label that risks flattening its complexity. The entire chapter is instead dedicated to analysing the combinatorial fabric among value, emotional and sense-motor layers, in order to bring out inferential patterns.

Although not a simple classifier, we provide here two experiments that test the functioning of our value detector, to show its performance with respect to various other approaches.

Therefore, to provide a baseline to demonstrate that the value detection is not randomic, we compare our tool with some baseline models. To have a meaningful comparison, we consider mainly non-trained models, as it would have neither sense nor purpose to compare it with deep learning or BERT-based models.

3.3.3 MFTC Experiment

This example has been published in [6] as a comparison for automatic value detection from natural language. We report here the experimental setting and some technical information about different models involved[15].

Through the reuse of ValueNet, this experiment solely focuses on Haidt's Moral Foundation Theory (MFT).

Starting from the method developed by [349], we adapt a checkpoint for BART-large[16] trained on the MultiNLI (MNLI) dataset [172]. Since this model has been shown to perform well for topic labeling [170] and for claim verification [265], it is a reasonable candidate for our task. Zero-shot learning methodology is a machine learning technique that enables a model to categorize new, undiscovered classes without having any training instances for them. In order to accomplish this, the model is trained on a set of classes for which there are labeled instances. At test time, the model is then given additional data, such as class attributes or class embeddings, regarding the unseen classes. The model can now transfer its understanding from the seen classes to the unseen classes and make predictions for them thanks to the added information.

Considering a Zero-Shot methodology, as first step we examine the input text for any concept similarities between its content and the moral values denoted by the labels. To the premise represented by the original textual data, we place side by side the categories suggested by Haidt's taxonomy as plausible hypotheses. In other words, we verify how much every value in the MFT's set is semantically related to every tweet in the test set (e.g. we evaluate if the concept *care* is expressed in the text 'Commitment to peace, healing and loving neighbors. Give us strength and patience.'). The same tweet is flanked by all the remaining moral values in the same way. The structure is based on the technique of using pre-trained NLI models as ready-made zero-shot sequence classifiers to develop a hypothesis from every possible label. As the output of the classification, results are acquired according to the predicted degree of entailment. The result of the categorization is represented by labels with a compliance score of 90% or above.

In the second step, we improve the model's prediction performances by adding more information on the latent emotional component in the original text.

To examine the effectiveness of our approaches in the moral value detection task, we focus on the challenge of recognizing them in the Moral Foundation Twitter Corpus (MFTC) [147].

To set performance baselines, we treat the annotations of the tweets by calculating the

[15] The results of this experiment come from different approaches, in particular the graph-based detector, es explained in Sec. 1.6.1 and 3.3.2 is developed by the author of this work, while the Zero-shot approach is developed by the co-authors Luana Bulla and Misael Mongiovì, whom we deeply thank for their contribution.

[16] `https://huggingface.co/facebook/bart-large-mnli`

majority vote for each moral value, where the majority is considered 50% (i.e. tweet 'I have no respect for the *home run king*' is labeled by four different annotators. Two of them regard the text as *Non-Moral* while the others as *Subversion*". Hence, we consider the tweet labeled as "Non-Moral, Subversion" because each of these labels corresponds to 50% of the annotation).

Table 1 shows the results obtained by our tools on a subset of 6,075 items representing the MFTC test set. Each tool is evaluated in terms of precision, recall and F1 score in predicting each label. The overall results (All in the bottom) are calculated by averaging over all labels weighted by the support (i.e. the number of elements in the ground truth with each specific label).

The Emotion-Zero-shot model displays the results obtained by exposing the Zero-shot model to an input text that has had its emotional component explained.

The Emotion-Zero-shot+ architecture refers to the combination of the two methods mentioned above.

The frame-based system recalls the results obtained from the application of the value detection pipeline described in Sec. 1.6.1. Given the lack of a reasonable state-of-the-art baseline of non-trained systems, we report a Random lower-bound, obtained by predicting each label with a probability corresponding to the fraction of entries in the ground truth represented by the test set with that label.

Finally, in Table 3.9 there is no reference to the *Liberty / Oppression* dyad. This happens coherently to the lack of this label in the MFTC, due to the late introduction of this value / violation opposition in an updated version of the MFT.

Triggers of this dyad are still detected by the frame-based model, and could be explored in the extended file [17], since the Liberty and Oppression knowledge graphs are part of ValueNet, but they are not considered in the evaluation metrics.

Furthermore, since the original dataset is annotated considering a 50% percentage of agreement among annotators, some of the sentences shows a combination composed by Non-Moral + some other value or violation. While for the Zero-shot models the Non-moral label is used as a feature itself, the combination of non-morality and any kind of morality is in conflict with the conceptual structure of the frame-based detector. We therefore modified the original dataset eliminating the Non-Moral label while co-occurring with some value or violation, and repeated the experiment.

Although performances differ, the two methods perform similarly in terms of F1, with an overall score of 45%. Specifically, Emotion-Zero-shot+ and Frame-based outperform the other models for four out of eleven labels, with F1 scores ranging from 0.12 to 0.53 for the first and from 0.11 to 0.50 for the second. These two architectures result in an

[17] https://github.com/StenDoipanni/MoralDilemmas

Moral Value	Metric	Random	Zero-shot	Emotion-Zero-shot	Emotion-Zero-shot+	Frame-based
Care	Precision	.09	.29	.51	.29	.29
	Recall	.18	.63	.36	.69	.57
	F1-score	.11	.40	**.42**	.41	.39
Harm	Precision	.13	.30	.31	.29	.39
	Recall	.24	.80	.59	.82	.70
	F1-score	.17	.44	.41	.43	**.50**
Purity	Precision	.04	.07	.10	.07	.18
	Recall	.08	.28	.30	.32	.20
	F1-score	.05	.11	.15	.12	**.19**
Degradation	Precision	.04	.12	.15	.12	.45
	Recall	.09	.63	.30	.66	.11
	F1-score	.06	**.20**	**.20**	**.20**	.18
Loyalty	Precision	.07	.40	.73	.40	.40
	Recall	.15	.45	.14	.46	.30
	F1-score	.10	.42	.24	**.43**	.34
Betrayal	Precision	.05	.17	.37	.17	.57
	Recall	.10	.44	.29	.44	.17
	F1-score	.07	.25	**.32**	.25	.27
Fairness	Precision	.07	.60	.85	.58	.16
	Recall	.15	.47	.26	.48	.11
	F1-score	.09	**.53**	.40	**.53**	.13
Cheating	Precision	.11	.54	.64	.54	.75
	Recall	.22	.29	.19	.29	.28
	F1-score	.15	.38	.30	.38	**.41**
Authority	Precision	.04	.17	.40	.18	.15
	Recall	.08	.28	.04	.29	.08
	F1-score	.05	.21	.07	**.22**	.11
Subversion	Precision	.08	.20	.15	.17	.28
	Recall	.16	.36	.39	.40	.17
	F1-score	.11	**.25**	.21	.24	.21
Non-moral	Precision	.44	.40	.46	.47	.59
	Recall	.66	.28	.86	.91	.72
	F1-score	.53	.33	.60	.62	**.65**
All	Precision	.22	.35	.46	.38	.47
	Recall	.36	.41	.52	.67	.48
	F1-score	.27	.35	.42	**.45**	.44

Precision, Recall and F1 score for each model on the MFTC dataset.

improvement of 10 % compared to the Emotion-zero-shot model and 20 % compared to the Zero-shot model, and they performs vastly better than Random.

Discussion As expected, the results for the single labels vary according to the difficulties encountered by classifiers in the interpretation of their meaning. For example, moral values such as *Harm* or *Care* convey more generic content and are therefore easier to identify. Conversely, concepts like *Degradation* or *Subversion* contain shades of meaning that are more difficult to grasp.

The results drawn from the Zero-shot models make this problem evident and difficult to solve as the intrinsic nature of machine learning models does not encompass a direct understanding of their decision-making phases. One possible solution would be to subject the models to few-shot learning, which is a fine-tuning with a little amount of data relevant for the moral values detection task. However, this would not be part of our main need, which is to develop flexible approaches that do not require training. Despite the task's complexity, the results imply that moral values can be detected in natural language texts.

Results obtained from the frame-based value detector are provided as additional material[18].

Value triggers are listed in the "trigger" column, while value detection is shown in the "prediction" column. The full knowledge graph can be retrieved by passing the tweet content in column "tweet_text" as input to the FRED online demo[19], ticking the "align to Framester" option.

A necessary caveat is that, being the value labeling a subjective task, a certain amount of disagreement should always be taken into account. In this regard, the detection shows better results on those values whose extension seems more generic, e.g. a more broad concept like *harm*, than a more opaque one like *sanctity*, as described in Sec. 3.2.4. Additionally, the performance results could depend on two factors. The first factor is the success of the FRED tool in producing a knowledge graph from a fragmented syntax like the one used in tweets. In fact, even when a well formed graph is produced, if the value trigger is not in the main sentence e.g. it is an adjective of a pronoun in a subordinate sentence, it is possible that its disambiguation/frame evocation is not shown in the graph, due to internal FRED saliency heuristics. The second factor is that human value labeling is a task carried out with a certain subjective threshold. If we consider the example: 'Horrible amount of anti-Islam bigotry are Paris attacks. ISIS murder more MUSLIMS than anyone else.', value labels for this sentence are "cheating" and "harm", while the detector predicts "cheating", "harm" and "sanctity". This happens because, along with triggers like the `fs:Offenses` and `fs:Killing` Framester frames, `wn:murder-noun-1`

[18] https://github.com/StenDoipanni/MoralDilemmas
[19] http://wit.istc.cnr.it/stlab-tools/fred/demo/

WordNet synset and the `db:Bigotry` DBpedia entity, the DBpedia entry `db:Muslim` is also retrieved, which according to `mft:Sanctity` definition (see Sec. 3.2.4) covers the semantics of a more spiritual aspect of life, and it is therefore a "sanctity" trigger.

3.3.4 MFRC Experiment

We conducted this experiment to test the ValueNet coverage and inference power. At the moment of writing, a smaller version of this experiment is included in a paper submitted to ESWC conference, still under review.

In order to perform automatic detection of values from natural language, we apply again in this experiment the value detection method, as described in Sec. 1.6.1.

To test the resource we selected from the Moral Foundation Reddit Corpus [327] (MFRC) a subset of the first 1k sentences.

In this experiment we tested our detector against four possible baselines, considering that they are different methods and that cannot be directly compared: (1) a random majority class algorithm, (2) a random classifier, (3) a decision tree algorithm and (4) a Zero-shot model as described in Sec. 3.3.3. Here a brief description of pros and cons, and a quick explanation of how these different methods work:

1. Random majority class algorithm: as the name could suggest, this is a very naive approach. It is not based on understanding the data, but purely on predicting always the majority class in the training data. It is therefore needless to mention that this method is neither a semantically relevant one, nor interesting from a feature extraction point of view, but it is even semantically highly misleading, as we will see in the following;

2. Random classifier: this method consist in randomly assign labels to the to-be-classified data, without any consideration for features or statistic relevance. The pros of this algorithm is that, being totally random, it can be used as baseline for any model;

3. Decision Tree algorithm: it is a model that, after converting text data into numerical features, create a decision tree classifier, trained using the above mentioned numerical features of the training dataset (and the corresponding labels). It is a baseline that, while being a (naive) trained method constitutes a possible baseline for our knowledge base model, albeit the two have a different approach to solve the same problem.

Table 3.5 shows quantitative data about the corpus used in this experiment: out of the 1k subset, shown in column "MFRC", in 944 cases, column "FRED Subset", the FRED

Table 3.4: Total amount of sentences per annotator, agreement

Annotators	Tot	Tot-NC	Agree / TOT	Agree+TM / Tot	Agree+TM/ Tot-NC
A00	157	63	52	62	34
A01	137	136	53	60	60
A02	185	180	65	75	75
A03	302	296	122	130	130
A04	163	163	6	63	63

tool successfully generated a knowledge graph. The reasons for the missing sentences not producing any knowledge graph could be due to many different problems: irregular syntax, brevity of sentences, use of abbreviation or not-recognised slang (e.g. “imho” for “in my humble opinion”, etc.), or even problems in character encoding.

Table 3.5: Total sentences and Original Annotation vs Frame-based Detected Values

MFRC	FRED Subset	MFRC Annotation	TM	NM	Detected
1000	944	228 / 1000	153	563	855 / 944

Table 3.4 shows some results of the analysis, in fact, the 944 sentences are not a set of unique sentences, this would be of 306 sentences, since the original corpus was realised using 5 annotators, granting each sentence to be labeled by at least 3 annotators. Therefore, each sentence annotated by each annotator is considered a token *per se*: in Table 3.4 column “Tot” shows the amount of sentences annotated by each annotator in the considered subset. The original dataset included also a confidence score, expressed as “Confident”, “Somewhat Confident”, and “Not Confident”. Column “Tot-NC” shows the amount of sentences per each annotator if we exclude those for which the confidence score is equal to “Not Confident”. Taking into account that this is a subset, it is still worth noting the uncertainty and intrinsic subjectivity of value annotation task: as Table 3.4 shows, A00 seems to be not confident almost half of the time, while A03 expresses confidence in 98% of the annotations.

A surprising datum is shown both in Fig. 3.9 and Table 3.5: out of 944 sentences, in the original annotation only 228 of them were tagged with at least one MFT value, as shown in Table 3.5, column “MFRC Annotation”. In fact, in 716 cases the sentence was consider not containing any MFT value, and therefore labeled as “Non-Moral” in 563 occurrences, shown in column “NM”, or having a “Thin Morality” in 153 occurrences shown in column “TM”. For the purpose of our analysis, we could paraphrase this datum

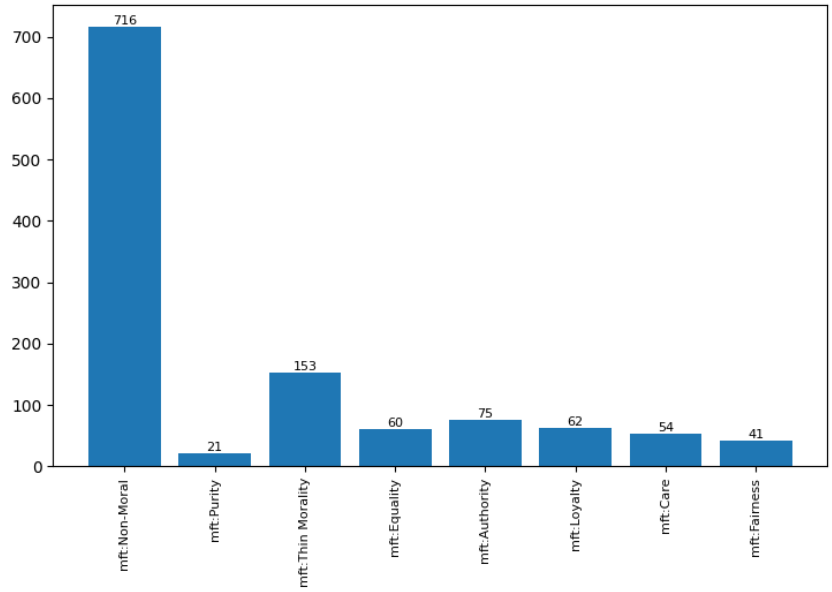


Figure 3.9: Original annotation of the MFRC dataset.

as: in 563 cases the MFT values were not retrieved in the sentence, while in 153 cases the annotator recognises that there is some sort of morality, but the MFT values are not enough to catch that specific/more subtle/cultural-dependent morality shade that the annotator still is claiming to be there.

The following paragraphs analyse all the baseline models' results and provide a details about their interpretation. For each of the following tables the column "Annotator" indicates the ID of the original MFRC annotator; column "Absolute" indicate the accuracy score of the model, considering the whole 1k corpus subset; column "Moral only" instead, shows the accuracy score for the model, considering the 1k subset excluding cases for which the annotator's label is "Non-Moral" or "Thin morality". The reason for doing so is twofold: (i) as stated in previous experiment, the "Non-Moral" label is given by the Graph-based if, and only if, no other value is retrieved, therefore it is not a real value-label, while for other classifiers instead it is interpreted as such, (ii) since, as shown in Tab. 3.5, the majority of the corpus is labeled as "Non-Moral", for the previous reason this would introduce a bias towards models that privilege more statistically relevant labels. Finally, class "Confident only", takes into account those cases for which the annotator expressed a confidence of "Confident" or at least "Somewhat Confident", and excludes those cases labeled as "Not Confident".

Majority Class As mentioned above, the majority class tends to reflect the statistical significance of the data, so it is not surprising that, as shown in Tab. 3.6, from a sizable 60% accuracy rate, it drops by about 20%, and even 40% in case of Annotaror04, when considering only those cases reported by annotators as containing some form of morality. There seems to be an understandable correlation between the attribution of morality and the confidence label per Annotator.

Table 3.6: Majority Class algorithm accuracy score.

Annotators	Absolute	Moral Only	Confident Only
A00	0.63	0.41	1
A01	0.66	0.31	0.65
A02	0.70	0.48	0.72
A03	0.79	0.39	0.81
A04	0.61	0.25	0.61

Random Classifier The random classifier considers the labels provided and then tries to guess at random, without considering any semantics. It is the only real baseline for our model, but, as shown in Table 3.7, results are so disappointing that it is clear that random attribution is not a methodology that can work in a natural language value annotation task.

Table 3.7: Random classifier algorithm accuracy score.

Annotators	Absolute	Moral Only	Confident Only
A00	0.05	0	1
A01	0.06	0	0.06
A02	0.17	0.11	0.13
A03	0.05	0	0.13
A04	0.05	0.07	0.08

Decision Tree The decision tree algorithm's accuracy, shown in Tab. 3.8, provides interesting results: although it shows respectable results in the "Absolute" column, these plummet dramatically when excluding cases noted as "Non-Moral." This is because, as mentioned earlier, decision tree algorithm is, although naive, a supervised learning method. As a result, since the sentences annotated as having no MFT values are the ones that provide the most material for feature extraction, in this case this result in the loss of over 40 percentage points in the "Moral only" column for both Annotator01 and Annotator04.

Table 3.8: Decision tree algorithm accuracy score.

Annotators	Absolute	Moral Only	Confident Only
A00	0.44	0.53	1
A01	0.43	0	0.53
A02	0.62	0.44	0.75
A03	0.74	0.36	0.68
A04	0.38	0	0.41

Zero Shot model Zero-shot model is the closest model to the approach proposed here: in fact, although its results are not excellent, it is the only model (excluding the graph-based one, described in the next paragraph) with any consistency. As shown in Tab. 3.9, there is almost no variation, apart from the "Confident only" column for Annotator00, but, as shown in Tab. 3.4, Annotator00 was not-confident half percent of the time, therefore it perfectly correlates with the 0.12 dropping to 0.06. This can be explained relatively to the fact that the pre-training of the zero-shot model on data that have nothing to do with value extraction, makes it an agnostic model with respect to the percentage of annotations per label (vs. majority class), specific features of the corpus under consideration, that would make the method too closely related to the corpus on which it is trained (vs. decision tree), and of course label assignment is not random, but defined by previously established/extracted parameters.

Table 3.9: Zero-shot model accuracy score.

Annotators	Absolute	Moral Only	Confident Only
A00	0.12	0.12	0.06
A01	0.10	0.10	0.10
A02	0.14	0.14	0.14
A03	0.22	0.22	0.22
A04	0.13	0.13	0.13

Graph-based model The graph-based model's results are shown in Tab. 3.10. As can be seen, not only is it the most consistent and coherent regardless of confidence and the presence of MFT values in the original annotation, but it also turns out to be overall the model with the best results among the nontrained.

Furthermore, other than MFT values, our detector is able to spot also BHV and Folk values, as described in Sec. 3.2.2 and 3.2.6.

Fig. 3.10, in particular, can be compared with Fig. 3.9, showing how, considering MFT only, there is a considerable increment in the variety, other than absolute number of

Table 3.10: Graph-based model accuracy score.

Annotators	Absolute	Moral Only	Confident Only
A00	0.39	0.39	0.50
A01	0.43	0.43	0.44
A02	0.40	0.40	0.41
A03	0.43	0.43	0.43
A04	0.38	0.38	0.38

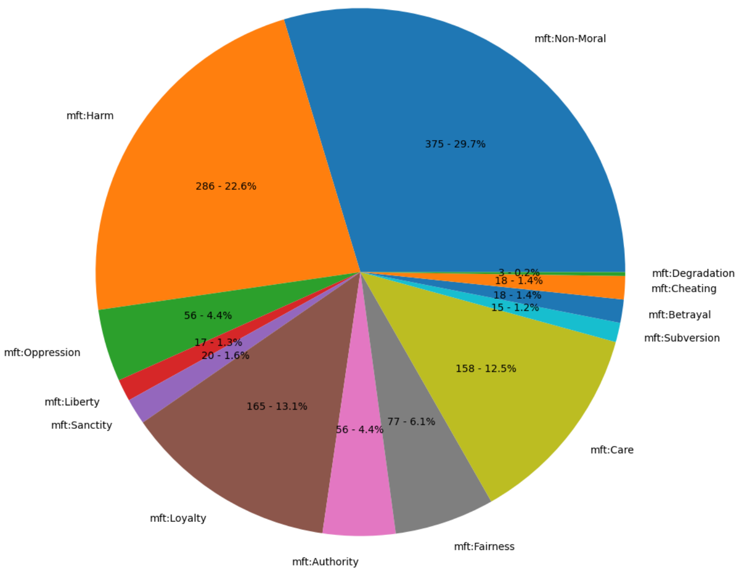


Figure 3.10: Moral Fonundations Theory detected from MFRC corpus subset.

occurrences of retrieval. The "Non-Moral" label is added in Fig. 3.10, in order to compare the amount of sentences for which no MFT value is detected from original annotation, and with our method. Note that, a higher amount of occurrences does not mean per se an improvement, but, as mentioned at the beginning of this section, they are necessary to understand the whole semantics of the sentence.

Fig. 3.11 shows the detection of BHV values: more than half of the BHV values detection occurrences consists in triggers for the `bhv:Power` value. The reason is probably that a consistent part of MFRC is composed by comments taken from subreds related to politics. Chapter 5 explores more in detail the co-occurrence of BHV, MFT and Folk values.

Considering instead Folk values, column "Detected" in Table 3.5 shows that, out of 944 total cases, in 855 of them at least one Folk Value is detected, 635 of which overlaps the

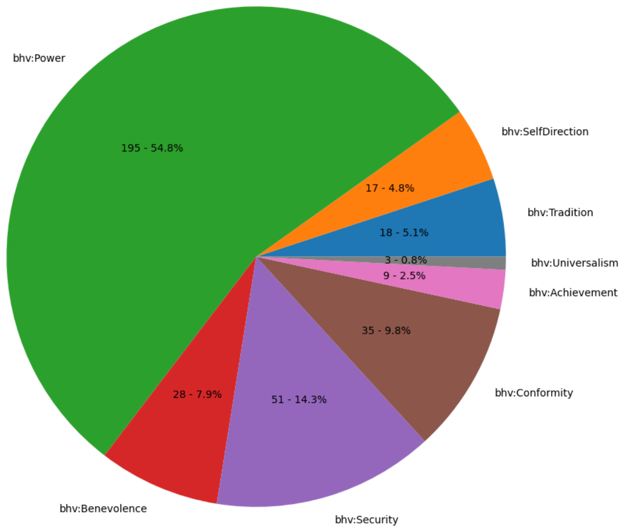


Figure 3.11: Basic Human Values detected from MFRC corpus subset.

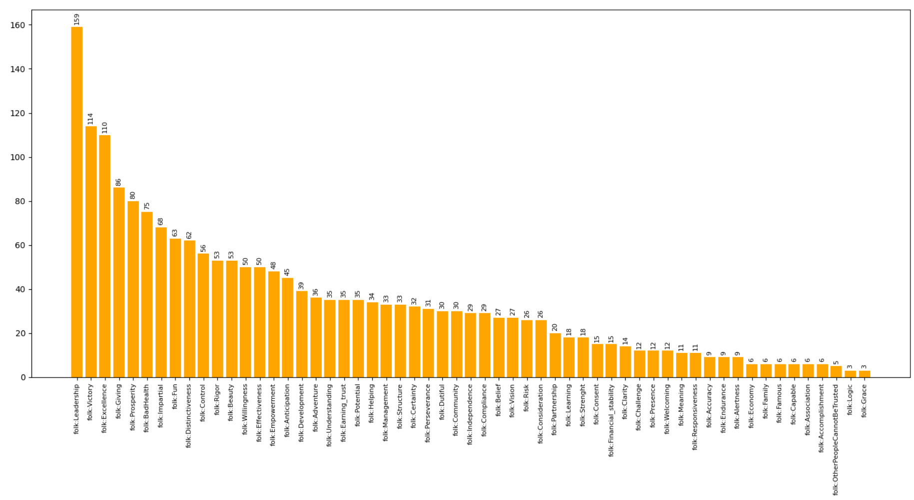


Figure 3.12: Folk Values detected from MFRC corpus subset.

subset of 716 cases for which zero or not-specified morality was originally indicated, Fig. 3.12 shows the detailed amount of activation occurrences per each Folk value. This alone, as clearly shown in Fig. 3.12, means a significant increment of the semantic information about latent moral content. To proceed with a more qualitative analysis: it seems that the subset that it is meaningful to be analysed here is the one composed by sentences that were labeled as having a “Thin Morality”, for which we can assume that the MFT values were not sufficient.

Let us therefore take an example, in order to show the automatic inferences allowed by the graph structure and the semantic dependencies. Graph n° 357 is generated out of the following sentence[20]:

> *And however flawed or dishonest Macron may be.....it is a far greater act of dishonesty to steal his data and expose it, hoping to change the course of a national election for the purpose of an outside group. That is far far more dangerous than voting for one flawed man.*

The sentence above is labeled by Annotator00 and Annotator03 as `Thin Morality` while it is `Non-moral` for Annotator01. The frame-based detector annotates this sentence with: `mft:Loyalty` and `mft:Betrayal` from MFT, and `folk:Rigor`, `folk:Learning` and `folk:Risk` from the Folk module. The full graph is not shown here for visualization reasons but it is available on the GitHub[21]. What is relevant is that, with the whole structure of semantic dependencies, it is possible to have the co-occurrence of some apparently conflictual tags, e.g. in this case the activation of both `mft:Loyalty` and `mft:Betrayal`. Analysing the graph, in fact, it is possible to retrieve the exact topology of activation, keeping track of the role of the value-trigger in the value situation. In this case `mft:Loyalty` stems from the `fs:Candidness` FrameNet frame, evoked by the “dishonest” lexical unit, and the WordNet synset for the adjective `wn:national-adjective-1`. The `mft:Betrayal` value is instead triggered by the `fs:RevealSecret` FrameNet frame, which is evoked by the VerbNet entity `vb:Expose_48012000`. As for folk values, `folk:Learning` is unfortunately activated by an incorrect disambiguation of the lexical unit “course”, creating a bit of noise; `folk:Rigor` is instead triggered by the `fs:Law` FrameNet frame, evoked by the segment “act of dishonesty” and `folk:Risk` is triggered by the `wn:dangerous-adjective-1` WordNet synset and the `fs:RiskySituation` frame.

Furthermore, via some graph-pattern heuristics, combining the VerbNet roles and the

[20] All the graphs generated by FRED and labeled with the value detector are available on the ValueNet repository: `https://github.com/StenDoipanni/ValueNet/tree/main/ThatsAllFolks/MFRC_1k_graphs`

[21] The full graph is available here: `https://github.com/StenDoipanni/ValueNet/blob/main/ThatsAllFolks/eswc_thin_folk.png`

affect stance of the verbs (available in the Framester resource), it is furthermore possible to retrieve, via SPARQL query, knowledge about the subject of the VerbNet entity `vn:Steal_10050000`, which is modeled as having a "socially reprehensible" negative value on the Agent (in our case the *dishonesty* node).

In conclusion, we are able to extract much and much more varied knowledge about the distribution of the sentence's value load, expliciting latent moral content, and offer an explainability that in a flat table would seem an inconsistency.

3.3.5 Chapter Conclusions

This chapter focused on the ValueNet ontology and the formalization of the main theoretical frameworks, namely MFT, BHV, Folk, and MM. Added to these are the semantic triggers modules of the value frames. Finally, two experiments were carried out, using an automatic frame-based value detector, one on the Moral Foundations Twitter Corpus (MFTC) and one on a subset of the Moral Foundations Reddit Corpus (MFRC) showing that using multiple theories at the same time can greatly increase the amount of value information automatically detected by natural language. The next chapter is focused on the emotion layer, on moral emotions and how do they correlate with the value layer.

Chapter 4

Emotions as a Layer of Knowledge

This chapter provides an introductory overview of the concept of "Emotion," while referencing some of the key theoretical frameworks. While the EmoNet ontological module is designed as a collection of theoretical frameworks that represent emotions, this work proposes it as a supporting module for investigating value knowledge.

4.1 Emotion Theory

Emotions, as well as values, are an extremely vast and varied domain that has been investigated by a wide variety of sciences, from philosophy [60], neuroscience [244], economics [205], biology [44], developmental psychology [248], sociology [330], and cultural anthropology [207]. Emotion and cognition have a long history of intertwined dependencies, often adopting a top-down approach (from a more philosophical perspective) while bottom-up approaches adopted a more empirical stance. Its application in computer science began as automatic extraction, first of mere positive or negative polarity, with sentiment analysis, and later with emotion detection techniques.

The first problem arises from the very definition of "Emotion" [257]. In fact, it can be viewed as a complex object and analyzed in terms of its mereology. However, when describing emotions, emotion theories primarily focus on the partonomic relations of emotion situations. The term "emotion" encompasses various aspects, including the triggering stimulus, the emotional state, immediate temporally consequent factors (e.g., facial expression), and the reactions (e.g., verbal or physical manifestations of appraisal) that an emotion entails [54].

The attempt to outline an emotion semantic frame from an incomplete list of emotion components would encompass the following elements:

- A cognitive component, involving various mental processes mediated by symbolic representation.

- A feeling component, representing the direct emotional experience and encompassing internal phenomenal or conscious aspects, which may also involve intentionality.
- A motivational component, consisting of action tendencies or states of action readiness, such as the inclination to fight or flee.
- A somatic component, indicating the somatic manifestation resulting from a neuro-physiological stimulus.
- A motor component, which may encompass more intricate behaviors, including facial and vocal expressions, as well as specific movement patterns.

Moreover, there is substantial disagreement among emotion theories regarding not only the precise number and nature of emotions but also the number of components involved in an emotional episode. This discrepancy can be expressed within the framework of Frame semantics as semantic roles participating in the Emotion Frame.

Moreover, there exists substantial disagreement among emotion theories not only regarding the precise number and nature of emotions but also the number of components involved in an emotional episode. This discrepancy can be expressed within the framework of Frame semantics as semantic roles participating in the Emotion Frame.

Furthermore, numerous theories compete to explain the role, dynamics, components, number, and even ontological status of emotions. Some influential works from the 19th and 20th centuries that have contributed to the flourishing research in this domain are as follows:

- The James-Lange theory: Named after William James [156] and Carl Lange [193], this psychological theory suggests that emotions arise from physical arousal caused by specific stimuli [36]. According to this theory, a stimulus such as physical contact, a noise, or an interaction triggers a neuro-physiological response in the body, leading to physical manifestations such as an increased heart rate, sweating, changes in body temperature, and breathing. The brain then interprets this physical response as a particular emotion. While this theory has been largely disproven by contemporary research, it has influenced subsequent theories of emotion, such as the Cannon-Bard theory and the Schachter-Singer theory.
- The Cannon-Bard theory: This psychological theory proposes that emotions and physical arousal occur simultaneously and independently of each other [64]. According to this framework, a stimulus would lead the brain to interpret a situation as, for instance, fearful while simultaneously inducing physical manifestations of the emotion. Although considered more accurate than the James-Lange theory, it

has been surpassed by more recent theories like the Schachter-Singer (two-factor) theory.

- The Schachter-Singer theory: Also known as the two-factor theory [282, 249], this psychological theory posits that emotions result from the interaction between physical arousal and cognitive interpretation. According to this theory, a stimulus triggers a physical response in the body, but the specific emotion experienced (e.g., enjoyment, anger, fear, etc.) is determined by the individual's cognitive interpretation of the arousal. This theory highlights the role of cognitive processes, including meaning attribution and perception. It continues to be widely considered and debated in contemporary research.

In addition to these theories, several emotion frameworks have been developed to aid in the understanding and categorization of emotions. Numerous theories have been constructed to explain theoretical assumptions and experimental data related to the concept of emotions. Here, we present the main ones and proceed to describe how some of them have been formalized. It is useful to differentiate models into two primary categories: (i) Categorical models and (ii) Dimensional models. Categorical models, including appraisal theories, organize emotions as discrete categories, often incorporating some form of taxonomy. These models primarily focus on a select group of concepts commonly referred to as "basic emotions" or "primary emotions." The number of basic emotions can vary depending on the categorical model adopted. For instance, Ekman's theory [68, 69] proposes five basic emotions and emphasizes the role of emotion facial expressions, serving as a foundation for tasks involving the automatic detection of emotions based on somatic manifestations. Mehrabian's theory [220] suggests six basic emotions, while Izard's theory [154] proposes seven. An alternative categorical approach, with a stronger linguistic orientation, is adopted by Shaver [301].

Regarding Appraisal theories, researchers such as Lazarus et al. [195], Ortony [238], Frijda [78], and others [275] describe the subjective process of assessing a situation that activates an emotional state in an individual.

One of the most prominent appraisal theories is the Ortony, Clore, and Collins model (OCC) [239], which posits that emotions are "valenced reactions to a process of appraisal" performed by individuals and influenced by various events, persons, or entities.

Dimensional models, on the other hand, can be further divided into vector models and circumplex models [30].

Russell's circumplex model [278] is one of the most well-known dimensional models. It positions emotions along a continuous space defined by two dimensions: arousal and polarity.

Another notable circumplex model is Plutchik's wheel of emotions. The Plutchik's wheel [250, 247] presents a taxonomic organization of emotions, with eight basic emotions (anger, fear, sadness, disgust, surprise, anticipation, trust, and joy), and "complex emotions" formed through combinations of adjacent emotions on the wheel.

In relation to appraisal theories, numerous studies have explored the connection between emotions and moral values. As mentioned in Chapter 3, social psychology studies by Greene and Haidt [114, 126], Rozin [277], and Scherer [201] have focused on moral emotions. These studies propose that emotions and morality are intertwined. Moral emotions, particularly from a socio-anthropological perspective, are considered coping mechanisms that aim to improve social relationships. They are seen as positive reinforcement nudges towards adopting cooperative strategies. The study of emotion and its connection to moral values is a multifaceted and continuously evolving field of research, with numerous theories and frameworks developed to elucidate this relationship.

For a concise historical overview of emotions, Evans [70] provides valuable insights.

To gain a deeper understanding of the theoretical frameworks used in this domain, De Houwer and Hermans [54] offer a comprehensive comparison.

For a comprehensive exploration of emotions and cognitive-related studies, we recommend consulting the Handbook of Cognition and Emotion [271] as well as the "Cognition and Emotion" journal, which provides a wealth of relevant research.

4.1.1 Emotion Dataset and Resources

We provide here a list of dataset and resources annotated with emotion content.

GoEmotions GoEmotions [59] is a comprehensive dataset annotated by humans. It consists of 58,000 Reddit English comments and includes 27 emotion labels along with a "Neutral" category. While it provides a taxonomy, it can be considered an annotative scheme rather than an ontological resource. The annotation process aimed to capture the labels that were most prevalent based on a bottom-up approach, disregarding major categorical frameworks such as Ekman, Plutchik, and OCC.

International Survey on Emotion Antecedents and Reactions (ISEAR) The ISEAR dataset [284] contains over 7,000 sentences comprising self-reported experiences from cross-cultural studies. It is annotated using the following basic emotions: joy, sadness, fear, anger, guilt, disgust, and shame.

DailyDialog The DailyDialog dataset [198], developed by Li et al., consists of approximately 1,000 dialogues. Each dialogue is annotated with emotions from the following dimensions: happiness, sadness, anger, disgust, fear, surprise, and neutral.

SemEval-2017 SemEval Task 4 [276] focused on sentiment analysis on Twitter. It includes a miscellaneous corpus of tweets, news headlines, Google News, and other news sources in English and Arabic languages. The dataset adopts Ekman's model, specifically the version with 6 basic emotions.

WASSA 2017 The WASSA 2017 corpus [227] was developed for measuring automatic detection of emotion intensities. It consists of four datasets, one per each emotion considered (joy, sadness, anger, and fear), of tweets annotated with with a single emotion label. Each dataset consists of approximately 1000 tweets, all labeled with one of the emotions. This dataset, unlike many others, is annotated on each single sentence, and not e.g. for an entire dialogue.

SMILE The SMILE dataset [341] is composed of 3085 tweets about the British Museum from May 2013 and June 2015. It uses Ekman's basic emotions: anger, disgust, happiness, surprise and sadness.

Emotion-Stimulus The Emotion-Stimulus dataset [102] includes approximately 1,500 sentences collected from Framenet. It focuses on both emotion frames and emotion cause triggers, resulting in a collection of emotion-labeled sentences. The annotation follows Ekman's basic emotion model, with the addition of "shame."

EmoBank EmoBank [33] is a resource based on a dimensional model. It consists of 10,000 annotated sentences from various types of documents, including news headlines, blog posts, essays, newspaper articles, and travel guides. The annotation is done according to the Valence-Arousal-Dominance emotion representation model.

DepecheMood DepecheMood [310] is a resource specifically designed for emotion detection tasks. It consists of a lexicon with approximately 37,000 terms, each annotated with emotion scores on dimensions such as afraid, amused, angry, annoyed, happy, inspired, sad, and "don't care." The lexicon is created through crowd-sourced affective annotation by scraping social media comments. Each term in the lexicon is associated with scores for each of the mentioned emotion labels. This resource is available on the Framester hub, aligned with WordNet synsets, allowing for the extraction of knowledge about emotional semantics directly from the Framester SPARQL endpoint.

WordNet Affect WordNet-Affect [313] is an extension of WordNet that includes emotion or valence annotations for synsets. The synsets are labeled with four labels: positive, negative, ambiguous, and neutral. One important aspect of this semantic schema is the stative/causative dimension. An adjective is considered "causative" if its semantics point to some emotion being caused by the entity described by the adjective (e.g., an amazing road trip). On the other hand, an adjective is considered "stative" if it refers to an emotion owned or felt by the subject described by the adjective (e.g., a caring or joyful person).

4.1.2 Emotion Detection

Emotion detection from text can be approached using various computational methods. Here are some of the techniques commonly used. Please note that this paper does not focus on emotion detection or utilize machine or deep learning techniques, so we will provide a high-level overview without delving into these topics.

Lexicon-based Approaches Lexicon-based techniques rely on utilizing one or more lexical resources to identify emotions. This approach can be further categorized into three subcategories: rule-based, ontology-based, and statistical.

1. Rule-based Approach: This approach involves encoding syntactical or logical rules to detect emotions from text. There are two common methods within this approach:
   - Keyword-based approach: A predetermined set of terms is used to classify the text into specific emotion categories.
   - Lexical affinity method: This method identifies relevant keywords and assigns probabilistic affinities to other words based on a reference emotion lexicon or model. This approach is often used for sentiment analysis, simplifying emotions to positive or negative polarities.
2. Ontology-based Techniques: These techniques leverage the relationships present in lexicon-based approaches. For example, EmotiNet [10] is a resource for detecting emotions from text that is built on a common-sense understanding of ideas, their interactions, and their emotional consequences. EmotiNet models situations presented in text as action chains based on appraisal models, considering the contextual information. It demonstrates that EmotiNet's structure and content are suitable for the automated treatment of implicitly conveyed emotions.
3. Statistical Approaches: Statistical techniques, such as Latent Semantic Analysis (LSA), aim to identify meaningful patterns among terms and sets of documents that

contain those terms. These methods can uncover hidden relationships and patterns associated with emotions.

For a more comprehensive survey on emotion detection, you can refer to the paper by Canales et al. [35].

4.1.3 Emotion Formalization

Although not in a large number, some ontologies have attempted to formalize the domain of emotions; they are briefly presented here. These ontological modules provide structured representations of emotions and related concepts, which can be used to enhance emotion detection, understanding, and expression in various applications.

EmOCA Emotion Ontology for Context Awareness (EmOCA) [21] is an ontological module specifically designed to improve emotion detection from physiological manifestations. It adopts Ekman's Basic Emotions theory as the basis for representing emotions.

EmotionsOnto EmotionsOnto [206] is an ontological module focused on emotion detection and expression systems. Its main purpose is not to represent human cognitive processes in emotional situations but rather to facilitate emotion detection and expression in computational systems.

SOCAM Ontology-based Affective Context Representation (SOCAM) is an ontology-based representation of affective states for context-aware applications. SOCAM allows for expressing the complex relations that exist among affective states and between these states and other contextual elements. However, please note that further details about SOCAM are not available at the moment.

MFOEM The MFOEM (Emotion Ontology) developed by Hastings et al. [133] is an ontological module dedicated to emotions within the OBO (Open Biological and Biomedical Ontologies) framework. It is based on the foundational ontology BFO [306] (Basic Formal Ontology) and incorporates the emotional theory expressed in Sanders and Scherer [279].

The MFOEM ontology provides a formal representation of affective phenomena and emotions, and it is designed for interdisciplinary use. Emotions in MFOEM are represented as `mfoem:emotion process`, which is a subclass of `mfoem:affective process`. Each specific emotion is listed as a subclass of `mfoem:emotion process`.

Here we list some relevant classes from the MFOEM ontology:

- `physiological process involved in an emotion`: This class represents the bodily processes associated with an emotion, including neurophysiological changes that occur in the central nervous system, neuroendocrine system, and autonomic nervous system. It serves as a superclass for various specific physiological manifestations of emotions, such as `mfoem:becoming pale` and `mfoem:heart beating at a faster rate`.

- `mfoem:mood process`: This sibling class to `mfoem:emotion process` represents mood-related processes. MFOEM distinguishes between moods and emotions, a distinction also present in the BE (Basic Emotions) ontological module.

- `mfoem:subjective affective feeling`: This class represents the valenced process of an emotion felt by an experiencer in response to internal or external stimuli.

- `mfoem:valence`: This class represents the valence, or the process profile, of an emotion. It can be positive or negative.

The MFOEM ontology provides a formal modeling framework for emotions and related concepts, allowing for a more structured and precise representation of affective phenomena in interdisciplinary research.

The MFOEM ontology stands out among the mentioned models for its high-quality modeling and alignment with a foundational ontology. However, in the current state of the EmoNet ontology work, MFOEM has not been included as a specific module, unlike the Ekman theory or the OCC. This is primarily due to the challenges posed by aligning two different foundational ontologies, namely DOLCE and BFO, which represent different philosophical views of the world. The debate between foundational ontologies has been ongoing, and this paper does not delve into the topic extensively.

However, future developments of EmoNet include plans to align and integrate MFOEM into the ontology. The goal is to bridge the gap and reconcile the two works.

Additionally, the framECO ontology, developed by Coppini et al. [41], is a transversal approach to representing emotion content in literary works. It adopts a frame structure and introduces specific classes to represent non-trivial emotion objects, drawing inspiration from the Dictionary of Obscure Sorrows, which catalogues "unfamiliar emotions." An example of such an emotion is "Onism," defined as the frustration of being stuck in just one body, limited to one place at a time, unable to experience the multitude of places represented on the departures screen at an airport. The framECO ontology is already integrated as a specific module within the EmoNet ontology.

In the next section we proceed to describe the modules for the main theoretical frameworks that have been transposed into ontological modules in EmoNet.

4.2 EmoNet Ontology

The EmoNet Ontology serves as an ontology network for representing emotions as frames, incorporating different emotion theories. By adopting a frame structure formalized within an ontological framework, it addresses the question of "what is an emotion" in its ontological sense. Emotions are considered cognitive objects or aspects of non-material reality, and according to specific theories, there are classes of entities that represent emotions within the context of those theories, reflecting different epistemic views of the world. Each theory's notion of emotion is represented as a frame that is a subclass of a more general emotion frame, with specific dimensions and aspects represented as roles.

Currently, the EmoNet Ontology consists of several modules. The EmoCore module encompasses the minimum vocabulary required to discuss emotions in terms of frames. The BE (Basic Emotions) module transposes Paul Ekman's theory, particularly the recent reworking found in the Atlas of Emotions, into ontological form. This module is described in more detail in Sec. 4.2.3. There is also the operationalization of the BE module, which involves populating the knowledge graph with Framester entities using the QUOKKA workflow. This workflow triggers one of the six basic emotions and is described in Sec. 4.2.4.

In summary, the EmoNet Ontology consists of multiple modules, including EmoCore, BE, and its operationalization, each addressing different aspects of representing emotions as frames within the ontological framework.

4.2.1 EmoCore

The EmoCore module serves as the core module within the EmoNet Ontology and provides the minimum vocabulary required to discuss emotions. It adopts a frame semantics approach to the emotion domain and models the concept of *emotion* as a frame. Aligned to the DOLCE foundational ontology, the EmoCore module utilizes the Description&Situation [95, 88] ontology design pattern to express emotions as both classes of situations and individuals (in OWL2 syntax).

Each emotion theory (already included or foreseen in future developments) is modeled in a separate module importinge the EmoCore one. The class `be:Emotion` is a subclass of `fschema:ConceptualFrame`, which in turn is a subclass of `dul:Description`. It is satisfied by instances of `emo:EmotionSituation`, which represents the occurrence or realization of a prototypical event involving an emotion. An emotion situation can encompass various aspects, such as a mental state, the expression of an appraisal consequence related to an emotion state, or the triggering moment of an emotion. As a core module, the EmoCore module generalizes specific notions of emotions to cover a wide range of

possible emotion situations. Its purpose is to provide a broad `emo:Emotion` class that can model all aspects of emotions as covered by different theories.
The EmoCore module can be explored online through its GitHub repository[1] or accessed via the Framester endpoint,[2] allowing users to query and interact with the ontology.

EmoCore Classes The EmoCore module includes several important classes that form the foundation of the ontology. The main class is `emo:Emotion`, which represents the broadest notion of an "emotion" within the EmoCore module. This class serves as the superclass for all theory-specific definitions of emotions. It is also related to two WordNet synsets, `wn:synset-emotion-noun-1` and `wn:synset-emotional_state-noun-1`, which provide synonyms associated with the concept of emotion.

The `emo:Emotion` class is a subclass of `fschema:ConceptualFrame`, which itself is a subclass of `dul:Description`. This hierarchy allows for the representation of emotions as frames within the ontology. Under the `dul:Situation` class, there are instances representing the realization of emotion situations, specifically the `fschema:FrameOccurrence`. These occurrences are further categorized into subclasses that align with existing FrameNet frames, providing different perspectives on the phenomenon of emotions. The subclasses include:

- `fs:EmotionActive`: This frame focuses on the positive or negative "push" that the Undergoer (the entity experiencing the emotion) feels from the emotion. It emphasizes the activity exerted by the emotion on a subject.

- `fs:EmotionDirected`: This frame describes an Experiencer who is feeling or experiencing a specific emotional response to a Stimulus or about a Topic. It captures the directed nature of emotions towards specific objects or situations.

- `fs:Feeling`: This frame represents an emotional state, which may involve an appraisal of the considered emotional state. It captures the subjective experience of an emotion.

- `fs:MentalProperty`: This general frame is used to refer to any possible mental state, including emotions. It can encompass both known mental states and those inferred from behavioral manifestations, such as psychological or physiological manifestations associated with emotions.

[1] The EmoCore module is available here:
`https://github.com/StenDoipanni/EmoNet/blob/main/EmoCore.ttl`

[2] The Framester endpoint is available here: `http://etna.istc.cnr.it/framester2/sparql`

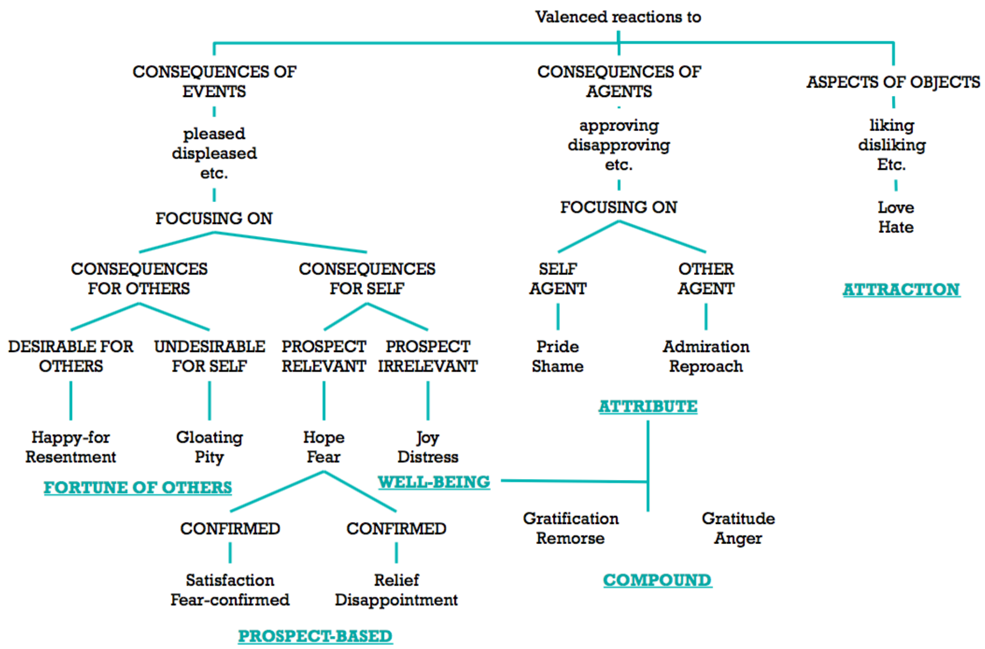


Figure 4.1: The Ortony Clore and Collins (OCC) model conceptual map.

These classes and frames within the EmoCore module provide a foundation for representing and understanding emotions in the ontology, capturing various aspects and perspectives of emotional experiences and states.

EmoCore Properties In the EmoCore module, the `emo:triggers` object property is a necessary property that is common to all different theoretical approaches and frameworks represented in the ontology. This property is used to declare that a particular entity serves as the trigger of an emotion.

The `emo:triggers` property allows for the association between an emotion instance and the entity or event that acts as its trigger. It represents the causal relationship between the trigger and the emotion, indicating that the presence or occurrence of the trigger leads to the experience or manifestation of the emotion.

By using the `emo:triggers` property, the EmoCore module enables the representation of the relationship between emotions and their triggers, providing a fundamental property for understanding and modeling emotions across different theoretical perspectives.

4.2.2 Ortony, Clore and Collins Appraisal Module Ontology

This module has been developed in collaboration with Rossana Damiano, to whom profound thanks are due for their contributions.

The OCC model is one of the main appraisal models in literature. Fig. 4.1 shows how emotions are taxonomically organised. In the following paragraphs we provide description of the main classes and properties.

OCC classes Fig. 4.1 illustrates the organizational structure of the original OCC model. Within this ontological module, the third branch, referred to as the "Aspects of Objects," is represented as an object property, which will be further described in the following paragraph. The remaining two branches of the graph, namely the "Consequences of Events" and "Consequences of Agents," are axiomatized in a symmetrical manner. In this discussion, we will focus on describing the "Consequences of Events" class, as depicted in Figure 4.1, which exhibits a greater depth in terms of the number of subclasses, thereby indicating a more stringent axiomatization. The "Consequences of Events" class utilizes the `dul:Process` class, while the "Consequences of Agents" class employs the sibling class `dul:Action`. Both of these classes, belonging to the DOLCE ontology, are subclasses of the overarching `dul:Event` class. Here, we present a list of the main classes and their respective axiomatizations.

- `occ:EventConsequenceEmotion`: as shown in Fig. 4.1, emotions as consequence of an appraisal process which takes as object some `dul:Process`. It is axiomased as follows:

  ```
  Equivalent To:
  feltBy some (AppraisingAgent and                (4.1)
  (appraises some Process))
  ```

  This axiom uses the `occ:AppraisingAgent` class, used to represent the subject of the appraisal performed by some cognitive agent, and distigueshed by another general class `occ:OtherAgent`;

- `occ:EventConsequenceEmotionOther`: emotions being the consequence of some action performed by some agent other than self. The action can be either praiseworthy or blameworthy for some appraising agent. It is axiomatised as follows:

  ```
  Equivalent To:
  feltBy some (AppraisinAgent and
  (appraises some (Process and
  ((blameworthy some AppraisingAgent) or          (4.2)
  (praiseworthy some AppraisingAgent)) and
  (performedBy some OtherAgent))))
  ```

  Subclasses of this class are those emotions respecting these restrictions, such as `occ:Gloating`, namely appraising some undesirable event happened to someone other than self, and finding it praiseworthy, therefore being happy of some undesirable event happening to other than self, or vice versa: `occ:Pity` being sad for some undesirable event happened to someone other than self.

- `occ:EventConsequenceEmotionSelf`: it represents emotions as consequences of some action performed by self. It is axiomatised as:

```
                         Equivalent To:
            feltBy some (AppraisinAgent and
                (appraises some (Action and
       ((blameworthy some AppraisingAgent) or
     (praiseworthy some AppraisingAgent)) and
        (performedBy some AppraisingAgent))))
```
(4.3)

  Its subclasses are furthermore restricted on the "prospect" dimension, i.e. being or not being dependent on the State s in which the agent is at a certain Time t. It subclasses are:

- `occ:ECES_ProspectIrrelevant`: emotions as consequence of some event happened to self for which the prospect is irrelevant, such as `occ:Distress` of `occ:Joy`. Its sibling class is:

- `occ:ECES_ProspectRelevant`: superclass for those dependent on some event localised at Time t+1 when the appraising agent is at Time t. These emotions are `occ:Fear` and `occ:Hope`. Finally, the subclass of this class is the one representing emotions for which the prospect was relevant, and it has been confirmed:

- `occ:ECES_ProspectRelevantConfirmed`: emotions like `occ:Disappointment` or `occ:Relief`.

- `occ:PolarityOutcome`: the polarity outcome of some emotion: positive or negative.

In the next paragraph we describe object properties, used in the OCC model to express the semantics for the classes above-mentioned.

OCC properties The OCC object properties are mainly focused on modeling the appraisal process, for its being positive, negative, praiseworthy or blameworthy. Here a list of the main object properties:

- `occ:appraises`: some Agent appraises some Entity;
- `occ:blameworthy` or `occ:praiseworthy`: some Entity is praiseworthy or blameworthy for some Agent;
- `occ:desirable` or `occ:undesirable`: some Situation is desirable or undesirable for some Agent;
- `occ:feels`: some Agent feels some Emotion. Its inverse is `occ:feltBy`;
- `occ:outcomePolarity`: some emotion has as outcome polarity a positive or negative polarity;
- `occ:performedBy`: some Process or Action is performed by some Agent.

4.2.3 Basic Emotions Module

The Basic Emotions module serves as an ontological representation of the Basic Emotions theory [68], as articulated and formalized on its dedicated website[3], with the inclusion of the *curiosity* emotion, which was not present in the original version. According to Ekman's theory, there exist six Basic Emotions: Enjoyment, Curiosity, Fear, Sadness, Anger, and Disgust. The current iteration of the theory encompasses additional aspects such as "Mood" and "Pre-condition," referring to internal or external states that influence the emergence of a specific emotion in response to a stimulus. The primary competency questions addressed within this module are as follows:

- **CQ1**: What and how many are the emotions included in Ekman's theory?
- **CQ2**: What is the polarity associated with each Basic Emotion?
- **CQ3**: Which psychopathologies predispose individuals towards specific emotions?
- **CQ4**: What are "emotion counters" or "emotion antidotes," and what counters/antidotes exist for specific emotions?
- **CQ5**: How does the intensity of one emotion compare to another?

BE Classes Here we list some of the main classes in the BE module:

- `be:PreCondition`: This class represents the context or situation that can influence how a subject experiences or enters into an emotion.

[3]The up to date Ekman's emotions theory can be found here: `www.atlasofemotions.org`

- be:BE_Emotion: The be:BE_Emotion class is used to represent entities that are considered emotions within the Basic Emotions theory. As primitive concepts, they are not provided with a formal definition. Each emotion takes as subclasses more specific states organized based on increasing intensity, as exemplified in the following turtle syntax:

```
be:Exasperation  rdfs:subClassOf  be:Anger  ;
        be:moreIntenseThan  be:Frustration  .      (4.4)
```

- be:EmotionCounter: This class represents a counterforce to an emotional state. If the emotion is positive, the counterforce is an be:EmotionImpediment, which refers to another emotional state that conflicts with the positive one. If the emotion is negative, the counterforce is an be:EmotionAntidote, which typically involves an intentional action or commitment. For example, the be:AnxietyAntidote is defined as 'Making a special effort to let go of ruminations about the past and anticipations of the future.'

- be:Message: This class represents the message conveyed by the emotion as a response to external stimuli.

- be:Mood: This class refers to longer-lasting emotional states that contribute to the repetition of the same emotional state even without an explicit trigger.

- be:PerceptionDatabase: This class represents the collection of universal or hardwired responses and individually acquired emotional memories. The contents of this database influence the appraisal process performed on a trigger and can therefore affect the resulting emotional state.

- be:Trigger: This class represents the interaction between the appraisal process and a hardwired or acquired script within the Perception Database.

- be:PersonalityTrait: This class refers to a tendency that predisposes a person to lean more frequently towards certain emotional states.

- be:PhysicalChange: This class represents the changes that occur in our body when an emotion arises.

- be:PhysiologicalChange: This class refers to the qualitative experience of an emotion, namely the manifestations that are determined by the emotion.

- `be:Psychopathology`: This class encompasses pathologies that can be attributed to certain emotions. Each pathology has a prototypical `be:emotionalTendencyTowards` a specific emotional state.

- `be:SelectiveFilterPeriod`: This class represents a state in which, upon encountering a certain initial trigger (e.g., fear), there is a narrowed and distorted perception, with filtering and interpretation of information that aligns with the prevailing emotion.

- `be:Signal`: This class refers to external and universally recognized prototypical manifestations of an emotion, which can be displayed through facial expressions or voice tone.

- `be:PostCondition`: This class represents the outcome of our emotional actions. The post-condition can be both external and internal, and it can influence subsequent emotional states.

BE Properties To represent semantic relations among classes involved in an emotion event, based on the Basic Emotions theory, the BE module includes the following main object properties:

- `be:emotionalTendencyTowards`: This property indicates that certain psychopathologies tend to incline individuals towards specific `be:BE_Emotions`;

- `be:hasAntidote`: This property signifies that a negative emotion is counteracted by an `be:EmotionAntidote`;

- `be:hasImpediment`: This property indicates that a positive emotion is impeded or inhibited by an `be:EmotionImpediment`;

- `be:hasPreCondition`: This property signifies that an emotional state has a `be:PreCondition`, which should be considered in determining the final `be:PostCondition` state;

- `be:moreIntenseThan` and `be:lessIntenseThan`: These properties indicate the relative intensity of one subclass of `be:BE_Emotion` compared to another subclass.

The Competency Questions presented in this section can be addressed by querying the BE ontological module. The following query provides answers to the BE competency questions by utilizing the classes and properties described in the preceding paragraphs, aiming to explore the ontological nature of emotions as modeled in Ekman's Basic Emotions theory.

```
SELECT DISTINCT ?emotion ?polarity ?psychopathology ?subEmotion
?antidote ?action
WHERE {
?emotion rdfs:subClassOf be:BE_Emotion .
?emotion be:hasPolarity ?polarity .
?psychopathology be:emotionalTendencyTowards ?emotion .
?emotion be:hasPersonalityTrait ?personalityTrait .
?subEmotion rfds:subClassOf ?emotion ;
be:hasAntidote|be:hasImpediment ?antidote ;
be:moreIntenseThan ?siblingEmotion .
?emotion be:hasAction ?action.
FILTER(regex(str(?emotion), 'Fear')) }
```

The query investigates the BE graph to retrieve all emotions that contain the term "fear" in their string representation. In this example, we focus on the be:Fear emotion. The initial triple pattern in the query requests entities (?emotion) that are subclasses of the be:BE_Emotion class. Among these entities, the only one containing "fear" in its string representation is be:Fear. Therefore, the query seeks information about the polarity of be:Fear, which is be:NegativePolarity, and identifies the psychopathologies whose patients tend towards be:Fear, namely be:AvoidantPersonalityDisorder, be:GeneralizedAnxietyDisorder, be:ObsessiveCompulsiveDisorder, be:PostTraumaticStressDisorder, and be:SocialAnxietyDisorder.

Furthermore, the query explores the personality trait associated with be:Fear, which is be:FearPersonality. It is described as 'A shy or timid person. This personality type is likely to avoid risks and uncomfortable situations. Timid people may perceive the world as full of difficult situations.' The subclasses of be:FearPersonality, such as be:Anxiety, be:Desperation, be:Dread, be:Horror, be:Nervousness, be:Panic, be:Terror, and be:Trepidation, are examined to identify if any of them have an associated antidote. The retrieved antidotes include be:AnxietyAntidote, be:DreadAntidote, be:HorrorAntidote, be:NervousnessAntidote, be:PanicAntidote, and be:TrepidationAntidote.

Finally, the query provides a list of sub-emotions for be:Fear, ordered by increasing intensity: Trepidation < Nervousness < Anxiety < Dread < Desperation < Panic < Horror < Terror. In this context, the symbol "<" indicates that one emotion is be:moreIntenseThan another.

4.2.4 Basic Emotions Triggers

This section is dedicated to the module that operationalizes the BE ontology. The KG population follows the QUOKKA frame building workflow, as described in Section 1.5. To illustrate and provide details about modeling choices, we consider the case of `be:Disgust`. However, the same process has been carried out for each `be:BE_Emotion`.

Additionally, Table 4.1 presents details regarding the number of entities retrieved as triggers from each resource for each emotion. The notable discrepancy between `be:Surprise` and all other emotions, particularly in terms of ConceptNet-aligned entities, can be attributed to the absence of `be:Surprise` among the basic emotions listed on `atlasofemotions.org`, the resource used for modeling the BE ontology. This discrepancy arises because Surprise is recognized as a universal facial expression but is debated as to whether it qualifies as a proper emotion. Despite this, it has been included in the set of BE emotions for two reasons: (i) it is still mentioned and addressed in the theory, albeit not as extensively as the other five Basic Emotions shown in Table 4.1; and (ii) many systems that perform automatic detection of emotions, based on Ekman's theory, utilize the six basic emotions as their dimensions, considering the original version of the BE theory, which is still one of the most cited and widely used. Therefore, it is deemed appropriate to include `be:Surprise` in the BE ontological module since it is well-documented in theoretical literature and considered in emotion detection experiments.

Frame Building Use Case: Disgust The knowledge graph for the Disgust frame triggers, based on the conceptualization and treatment of *Disgust* in the BE theory, is populated with numerous triggers from various resources.

The `be:Disgust` class actually has 7 subclasses, representing more specific types of disgust. These subclasses, listed in ascending order of intensity, are Dislike, Aversion, Distaste, Repugnance, Revulsion, Abhorrence, and Loathing[4]

These lexical units form the Starting Lexical Material (SLM) for the `be:Disgust` frame. Each unit serves as an input variable to retrieve the corresponding WordNet synset. For example, considering the unit "dislike," the following is a list of retrieved lexical units with synonymous and hyponymous relations: "dislike, disapproval, disfavour, disaffection, disinclination, unfriendliness, aversion, distaste, contempt, disdain, scorn, disgust."

It is worth noting that some of the retrieved units overlap with other `be:Disgust` subclasses. For instance, "aversion" and "distaste," which are retrieved in the same WordNet

[4]The symbol ">" is used here instead of repeating the `be:moreIntenseThan` object property multiple times.

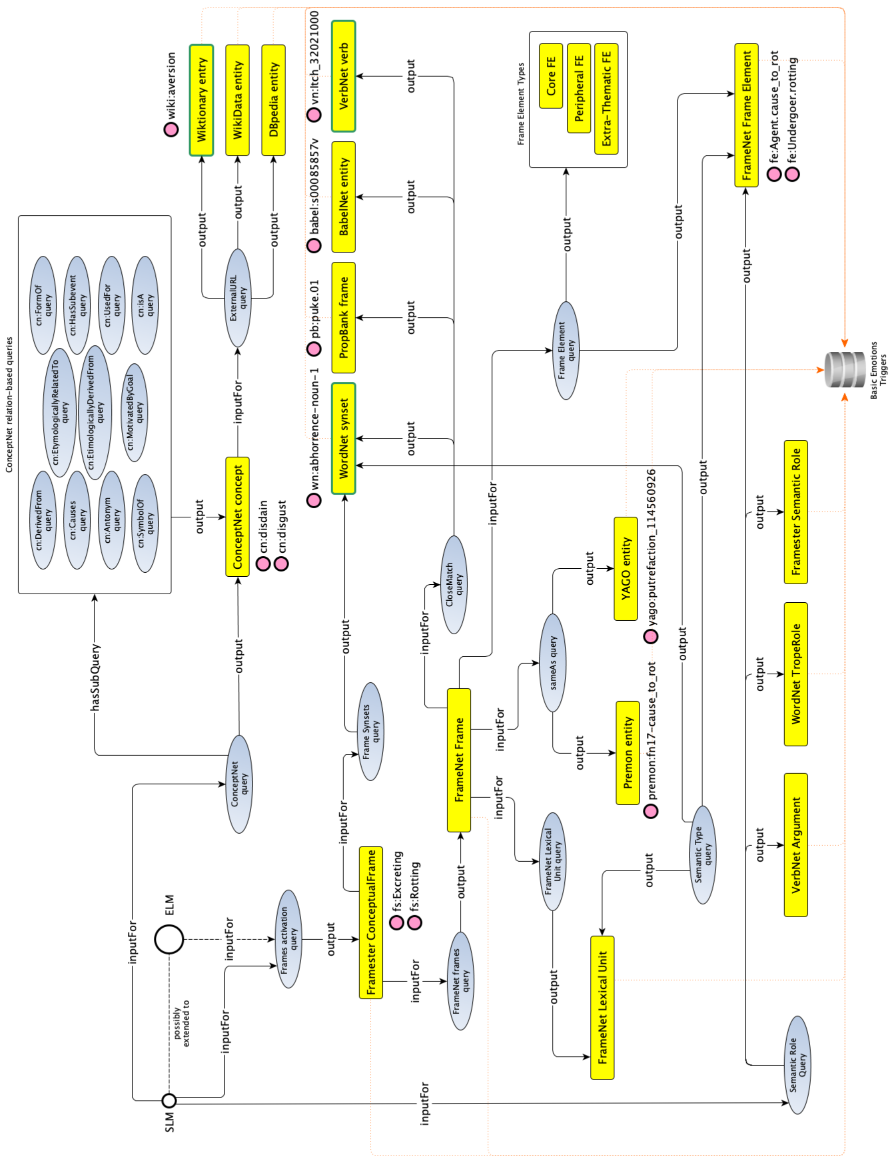


Figure 4.2: QUOKKA workflow applied to Basic Emotion knowledge graph population: be:Disgust triggers.

synset as "dislike," are also the names used for two other be:Disgust subclasses. This overlap indicates the existence of conceptual isoglosses, where terms refer to a certain aspect of meaning with blurred boundaries. These lexical pointers cannot be classified as triggering one and only one subclass of be:Disgust but rather demonstrate a fluid movement following an increasing intensity.

This finding reflects the highly debatable nature of the topic and is consistent with over forty years of disagreement regarding the nature, number, and structure of emotions.

Therefore, each unit in the aforementioned be:Disgust SLM set serves as an input variable for query expansion to retrieve entities from various semantic web resources in the Framester hub. Referring to Figure 4.2, the following observations can be made:

- Frames triggering: The frames fscore:Excreting, fscore:BeingRotted, fscore:CauseToRot, and fscore:Rotting are retrieved by the SLM set. These frames, manually selected from all those evoked by the SLM set, appear to be suitable triggers for the be:Disgust frame.

- Frame-element driven triggering: The be:Disgust frame inherits the previously mentioned frames as both triggers and its own frame elements. A Disgust situation occurrence is therefore triggered by elements such as fe:Manner.CauseToRot, fe:Undergoer.CauseToRot, and fe:Place.CauseToRot. However, Disgust as a frame inherits these elements, resulting in the automatic extraction of be:Disgust semantic roles, specifically be:Manner.Disgust, be:Undergoer.Disgust, and be:Place.Disgust.

- Lexical triggering: The be:Disgust frame is declared emo:triggeredBy all the entities subsumed by the previously mentioned frames, including WordNet synsets and aligned VerbNet entities. A be:Disgust situation is triggered, for example, by entities such as wn:synset-putrefactive-adjectivesatellite-1 and wn:synset-putrefy-verb-1, as well as their VerbNet alignment vn:Putrefy_45040000.

- skos:closeMatch triggering: The aforementioned frames connect to other entities declared to have a close match to them, allowing the declaration of triggers such as yago:rancidity_114561839, premon:fn17-excreting, pb:puke.01, and babel:s00028852n ("muck") from resources including YAGO, Premon, PropBank, and BabelNet, respectively.

- ConceptNet triggering: The SPARQL query expansion on the ConceptNet side, as depicted in Figure 4.2, is performed for each degree of intensity of be:Disgust, resulting in a significant number of triggers, such as cn:dislike and cn:disdain.

Finally, Table 4.1 shows details of each Basic Emotion frame.

4.3 Emotion Detection

To test the resource, we conducted automatic emotion detection using the WASSA 2017 corpus [227], which was generated in the context of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. This corpus was chosen because, unlike many others, it provides annotations for individual sentences rather than entire dialogues. However, a similar issue regarding the significance of certain annotations arises here as well. For instance, in an exchange of jokes between two individuals, the humor may not rely on explicit lexical triggers within a single sentence. At this stage of the detector's development, we are unable to make such complex inferences. The primary goal of these experiments was to test automatic detection and subsequently utilize it to support the exploration of value knowledge, which is why we operationalized only Ekman's Basic Emotions theory.

The WASSA 2017 corpus consists of four datasets, each annotated with one of Ekman's basic emotions: fear, anger, enjoyment, and sadness. Each dataset comprises approximately 800-1,000 tweets, all expressing only one of the aforementioned emotions according to the gold annotation. Although a considerable portion of these tweets, particularly the first few hundred, are clearly related to the respective emotion in various ways, the quality of the dataset cannot be considered optimal. This is because the tweets were collected based on chosen hashtags. Therefore, during the file preprocessing, we decided to remove the '#' character but retain the hashtag string itself. Many of the tweets consisted of only a few words, and the entire meaning depended on the hashtag as a semantic qualifier and a disclaimer for both the topic and possible senses of the sentence. Detection specifications are provided in the following paragraphs. The detection performed on this dataset also includes image schemas and values, and their co-occurrences are briefly analyzed, with a particular focus on the `be:Fear` emotion. The next chapter is entirely dedicated to exploring graph patterns that involve the co-location of image schemas, emotions, and values.

All the graphs generated during the graph-based emotion detection process are available in the EmoNet GitHub repository[5].

Enjoyment According to Ekman's emotion theory, particularly as axiomatized in the BE module, enjoyment encompasses various notions, such as `be:SensoryPleasure`, `be:Amusement`, and so on, leading up to its peak, `be:Ecstasy`.

[5]The repository can be accessed here: `https://github.com/StenDoipanni/EmoNet`

Emotion	Frame	Frame Element	WordNet	VerbNet	PropBank	ConceptNet	Wiktionary	WikiData	DBpedia	BabelNet	Umbel	YAGO	Premon
Disgust	4	37	227	69	22	418	341	10	10	78	0	29	4
Fear	2	21	152	40	0	199	130	6	6	46	4	35	2
Anger	2	28	309	78	12	304	229	7	7	102	2	40	2
Sadness	2	26	241	79	13	573	432	12	12	49	7	20	2
Surprise	1	14	51	21	0	42	27	2	2	11	1	8	1
Enjoyment	2	26	268	118	6	1131	900	48	41	26	53	19	2

Table 4.1: BE Emotions and amount of triggers from each semantic web resource involved in the emotion frame building process.

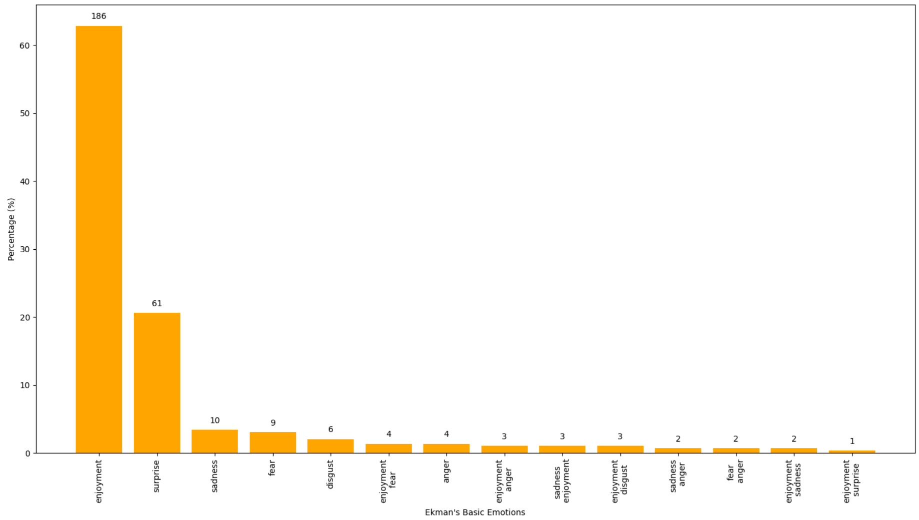


Figure 4.3: Enjoyment Basic Emotion automatically detected in WASSA 2017 corpus.

The Enjoyment WASSA2017 file contains a total of 824 tweets, all of which are labeled as expressing some form of enjoyment. Out of these 824 tweets, FRED was able to successfully generate a graph for only 711 cases, due to reasons such as their brevity, syntax, and others mentioned in Section 1.6.1.

Among these 711 tweets, the emotion `be:Enjoyment` is retrieved in 204 occurrences, `be:Surprise` in 67 occurrences, and combinations of various emotions appear in less than 10 occurrences each. Most of the sentences are not identified as having any emotion content, which can be attributed to the reasons mentioned earlier.

Considering Graph 676 as an example, it is generated from the sentence 'sparkling or still? #terribledebatequestions.' The tone of the tweet is humorous, and it is labeled with "joy." However, there are several issues with this sentence: (i) there is no direct lexical trigger for any emotion; (ii) even if we consider "terrible" in the hashtag as a trigger, it would be associated with `be:Fear` or `mft:Harm` rather than joy; (iii) the object is omitted since "water" is never mentioned; (iv) "terrible" is used ironically, and this can only be understood with commonsense knowledge, as the debate topic of 'sparkling or still' is not typically a matter of international politics.

It is important to emphasize that this emotion detector is a by-product of the investigation into the link between values and emotions. Therefore, it is more likely to provide incidental information rather than being used independently. Nevertheless, since the EmoNet ontology aims to be an ontological atlas of emotion theories, it is noteworthy that the detection does not exhibit a high rate of confusion. The most frequently detected emotion is `be:Enjoyment`, followed by `be:Surprise`, which are often strongly correlated through

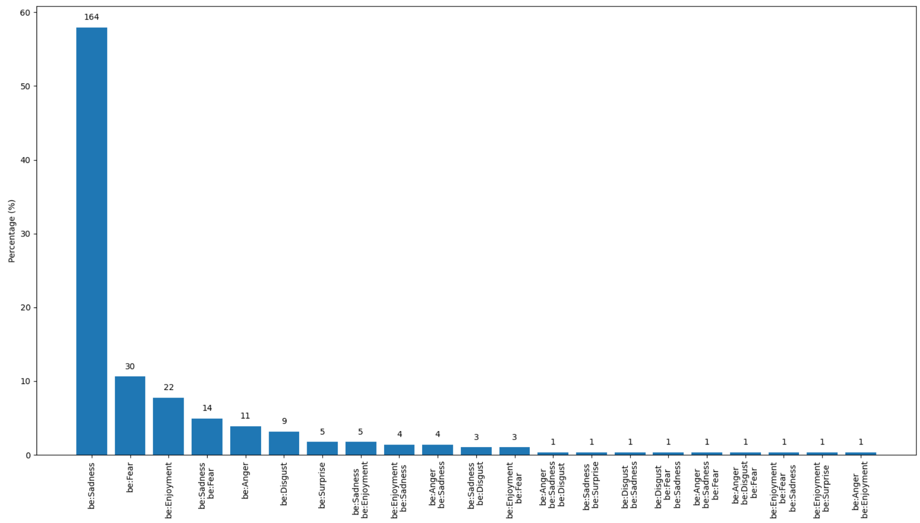


Figure 4.4: Sadness Basic Emotion automatically detected in WASSA 2017 corpus.

expressions such as “amazing,” etc. Occurrences of negative emotions, such as `be:Fear`, can be attributed to dialogic forms, as seen in Graph 266: ’At the risk of being blocked, I thought that the video from the wake was hilarious,’ where a `fs:RiskySituation` is evoked, triggering fear.

In general, to better identify occurrences of positive emotions, it may be fruitful to focus on values like `mft:Care` and `mft:Liberty` or image schemas like Support, and examine their role as the Patient, which could be the Experiencer of enjoyment.

Sadness The sadness WASSA2017 file consists of 786 tweets, out of which FRED generates graphs for 752 cases. Among these cases, some evocations of `be:Sadness` are retrieved in 200 instances. In this particular case, we provide an example where a single label, established a priori in the case of WASSA2017 (as none of the tweets are meant to have more than one label), is particularly inappropriate.

Graph 501 is generated from the sentence ’The immense importance of football is sometimes scary. When you don’t win, you are responsible for so many unhappy people - Arsene Wenger.’ For this sentence, the detector retrieves both `be:Sadness` and `be:Fear`. Both emotions seem applicable when considering the emotions that are crucial for understanding the meaning of the sentence. However, neither emotion appears to reflect an explicit intention or manifestation of fear or sadness. If a choice must be made, the aspect of `be:Fear` could be considered to be more dominant.

In contrast to the detection of `be:Enjoyment`, a potential pattern for better detecting `be:Sadness` situations could involve searching for negations of certain `mft:Care` trig-

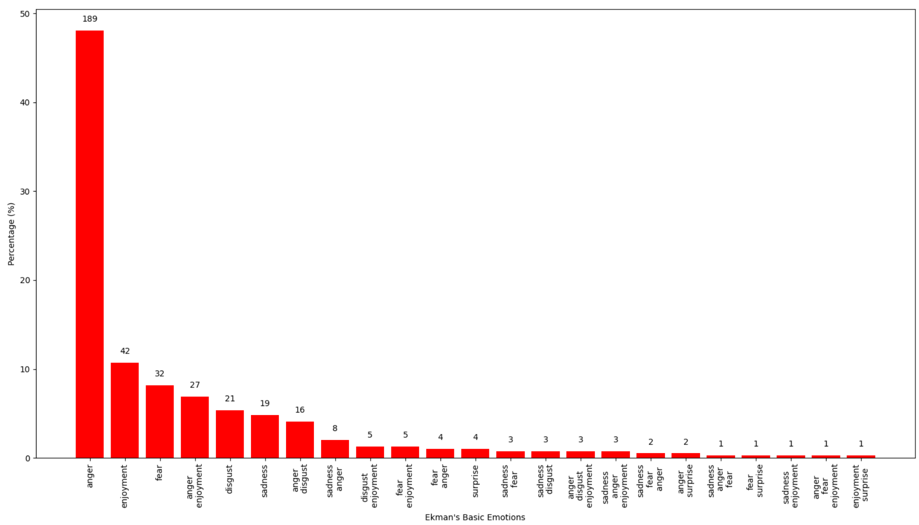


Figure 4.5: Anger Basic Emotion automatically detected in WASSA 2017 corpus.

gers. If the Patient of the negated `mft:Care` situation is a sentient being and evokes the `fs:People` frame, it could be inferred that the entity is experiencing some negative emotion.

A final observation is that, unlike values and image schemas, it appears that the emotional semantics are often conveyed through syntactically expressed adjectives functioning as qualifiers.

Anger The anger WASSA 2017 file contains 857 tweets, resulting in 816 generated graphs, with a total of 282 occurrences of `be:Anger`.

Figure 4.5 displays the specific number of occurrences. The significant number of `be:Enjoyment` evocations appears to be related to sentences such as the one in Graph 768: 'If someone keeps laughing at you, don't fret. At least you are giving happiness.' During the QUOKKA workflow for anger graph population, utilizing lexical units obtained from Ekman's theory available online, the entity `vb:Fret_31030800` was not initially retrieved as an anger trigger. However, after manual verification, it was subsequently included.

Furthermore, it appears that expressions related to `fs:EmotionHeat`[6] triggers, involving an Agent or Patient trigger associated with `wn:supersense-noun_body` and activating Containment, can be seen as a form of expression of the conceptual metaphor `THE BODY IS A CONTAINER FOR EMOTIONS`. This is particularly evident in situations involving the WordNet entity `wn:blood-noun-1`, and the VerbNet entities `vn:Fume_31030800`

[6] The `fs:EmotionHeat` frame represents the semantics associated with manifestations of emotions related to body temperature and includes situations typically associated with anger and love-arousal.

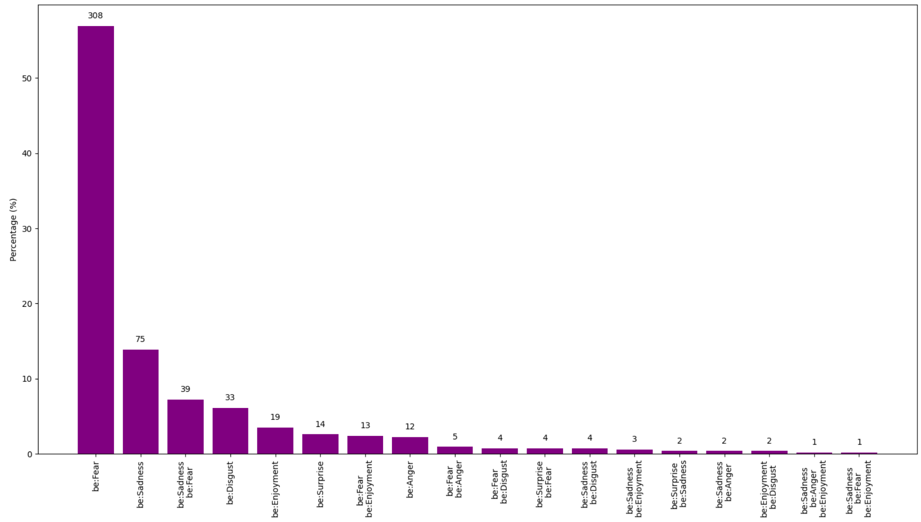


Figure 4.6: Fear Basic Emotion automatically detected in WASSA 2017 corpus.

and vn:Boil_45030000.

Interestingly, while no IS activation is retrieved for vn:Fume_31030800, the Agent trigger for vn:Boil_45030000 often corresponds to wn:blood-noun-1, which leads to wn:supersense-noun_body, consequently evoking Containment.

In abstracting and proposing a possible interpretation, it can be noted that although "boiling" and "fuming" are two ways of metaphorically expressing a state of anger, it is intriguing that in the former case there is an activation of Containment, as the idea of anger bubbling up in the body perceives the body itself as a container ready to explode due to the pressure exerted by the emotional state. Conversely, "fuming" indicates the same state of anger, but the focus is not on the pressure within the container; instead, it highlights the outward (involuntary) manifestation.

Fear The Fear file is the largest among the WASSA 2017 files, consisting of 1157 tweets. FRED generates graphs for 1076 cases, with be:Fear activation retrieved in 374 occurrences. The occurrences of other basic emotions and combinations are presented in Figure 4.6. There appears to be a noticeable correlation between be:Sadness and be:Fear, particularly concerning the concept of "despair." Interestingly, there are also 285 occurrences of mft:Harm retrieved.

We have measured whether our knowledge graph-based detection correlates with WASSA2017 annotations. Since WASSA2017 is a 'bronze' standard (labels have been derived automatically from Twitter hashtags, and manually evaluated on a small selection), the precision, recall and F1 measures reported in Table 4.2 are only indicative *as if* WASSA2017 were a gold standard. The Interrater agreement correlation measure is more

	Precision	Recall	F1 Score	Pearson Correlation
Anger	100	34.6	51.41	0.60
Fear	100	34.76	51.59	0.59
Sadness	100	26.23	42.06	0.64
Enjoyment	100	28.41	44.25	0.60

Table 4.2: WASSA 2017 Precision, Recall, F1 and Pearson Correlation score with frame-based emotion detector.

appropriate, because it compares two different approaches (WASSA2017 bronze standard and EFO knowledge graph-based detection).

Interrater agreement using Pearson r show moderate agreement, which may suggest a minor bias in the detection methods: hashtags can be motivated by content, opinion, or even irony; frame-based detection could be affected by contrasting frames. It could also suggest an inadequacy in the choice of the categorical model, or its labeling, since (as reported in [41]), real life affective situations can be much richer than what a simple categorisation could express.

The upcoming chapter will focus on the extraction of graph patterns and the analysis of co-occurrences between image schemas, emotions, and moral, social, individual, and cultural values.

Chapter 5

The Good, the Bad and the Container

This chapter focuses on conducting experiments that combine the three knowledge layers derived from the resources proposed in Chapters 2, 3, and 4. The initial section presents numerical data regarding the value detection carried out on the MFRC dataset using our automatic detection tool. Subsequent sections explore various ways in which the three knowledge layers (image-schematic, emotion, and value) interact, enabling the inference of new knowledge through graph-pattern exploration. The chapter demonstrates that graph-based value detection is more than a simple classification task; it leverages the graph structure to uncover sensorimotor, emotional, and moral knowledge embedded within natural language.

5.1 Graph-based Value Detection

The primary objective of this section is to describe instances of co-activation between different frames from the sensorimotor, emotion, and value domains, while identifying meaningful patterns of knowledge. This approach further emphasizes how, in the task of annotating sentence value content, a single label oversimplifies a vast amount of latent semantics, which our method not only reveals but also utilizes to trigger further inferences. The tool we propose is not intended to function solely as a classifier; rather, it serves as a value situation detector by automatically detecting moral atoms, as termed by Curry, to reveal the moral semantics of a given sentence. This, in turn, has three main outcomes: (i) it makes explicit the embodied conceptualization of a particular phenomenon by a cognitive agent; (ii) it elucidates the annotation process itself, by analyzing the semantic dependencies within the graph generated by a sentence; and (iii) it takes into account multiple layers of interdependent knowledge, not only pertaining to values, but also to the way in which moral issues are conceptualized (via the use of image-schematic expressions) and the emotional significance associated with each element of the sentence (through emotion detection).

To the best of our knowledge, no other work has attempted to exploit the graph structure by reusing semantic web resources for the detection of values, let alone integrating image schemas, moral/cultural/individual values, and emotions all together.

The subsequent sections are dedicated to the exploration of graph patterns. Specifically, Section 5.2.1 focuses on the occurrences of Image Schema (IS) situations, which involve a specific state of the world satisfying essential image-schematic configurations outlined in Chapter 2. The exploration of graph patterns concentrates on IS situations that incorporate knowledge from the value and emotion layers. This is achieved by utilizing the SPARQL query language to explore the graph structure and identify IS situations that encompass value or emotion triggers as semantic roles. The detection of semantic roles is facilitated by the FRED tool, as described in Section 1.6.1, with the semantic roles derived from the VerbNet resource. The examples presented in Section 5.2.1 showcase IS situation nodes where a value or emotion is assigned as a semantic role, such as Agent, Cause, Patient, and so on. This investigation also yields data regarding the co-occurrences of value, emotions, and IS, which hold relevance from a sensorimotor cognitive perspective. Section 5.2.3 revisits the investigation, focusing on Ekman's basic emotions. The SPARQL query is modified to search for emotion situations where a value is assigned as a semantic role. Lastly, Section 5.2.5 examines intricate conglomerates of interconnected value situations, wherein value situations take on the roles of other value triggers. Additionally, patterns of inference that combine different value theories are presented. The final discussion describes "commonsense knowledge patterns" derived from the investigations conducted in the preceding sections, which are then translated into the SPARQL language to enhance inference capabilities.

5.1.1 MFRC Automatic Image Schema, Emotion, and Value Detection

To test the hypotheses mentioned earlier and integrate different knowledge layers, we employ the frame-based methodology described in Section 1.6.1, utilizing the FRED tool to generate knowledge graphs from natural language. The extraction of Image Schemas (IS), values, and emotions, as performed in Chapters 2, 3, and 4, is repeated by unifying all the different detectors into a combined one.

The experiment conducted on a subset of the Moral Foundations Reddit Corpus (MFRC) in Section 3.3.4 is now expanded to a much larger set. The initial subset in Section 3.3.4 consisted of 1,000 non-unique sentences (approximately 300 unique sentences). However, for this experiment, we significantly increased the subset considering three factors:

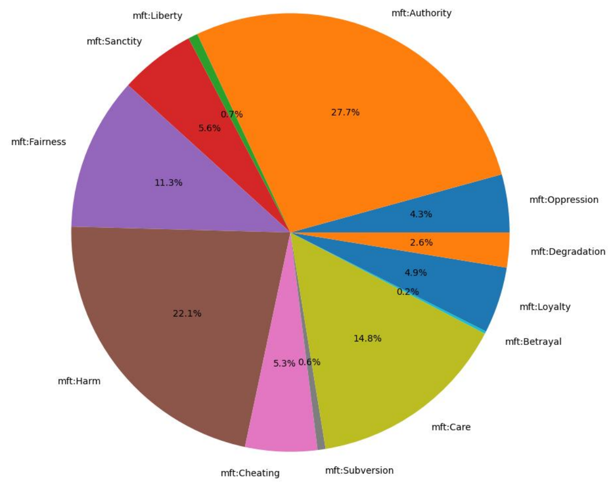


Figure 5.1: MFT values automatically detected in MFRC corpus.

1. The length of sentences in MFRC: Since Reddit comments have more elaborate sentence structures compared to tweets, the average length per comment is about two lines each;

2. The time-consuming nature of the automatic value-emotion-IS detection process: This process involves (i) knowledge graph generation by FRED and (ii) SPARQL queries to the Framester endpoint for each node retrieved from a semantic web resource from the Framester hub;

3. The size of the MFRC corpus: The total number of unique sentences is approximately 17,800, each annotated by at least three annotators, resulting in more than 61,000 differently annotated sentences. For our experiments, we considered the threshold of 75% of unique MFRC sentences as acceptable.

As a result, our final corpus consists of approximately 13,800 unique sentences. The entire annotated corpus, which includes the original information along with IS-emotion-value detection, is available on the ValueNet GitHub repository. Additionally, a zip folder containing all 13,800 generated knowledge graphs, used for graph pattern experiments, is provided in the same repository to facilitate in-depth graph exploration.

Here are some statistics regarding the percentage of value activation. From this point forward, unless otherwise specified, all diagrams and statistics refer to the subset of 13,800 unique MFRC sentences.

The MFT diagram depicted in Figure 5.1 confirms the findings presented in Table 3.2 in Chapter 3. The primary value frames evoked are `mft:Authority` and `mft:Harm`, with

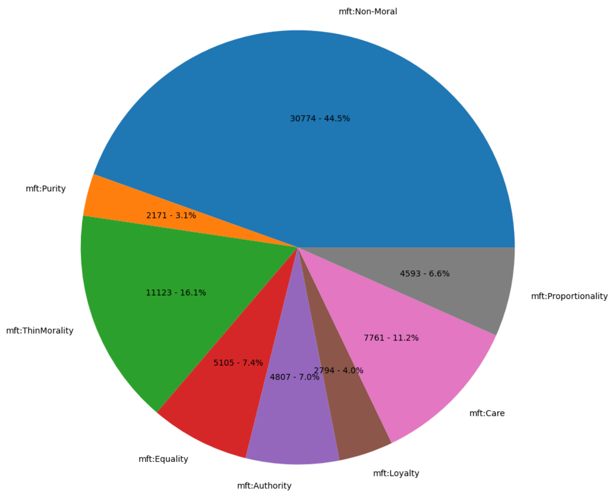


Figure 5.2: MFRC original annotation.

the former aligning well with the topics covered in the MFRC dataset. Specific violations such as `mft:Betrayal`, `mft:Subversion`, and `mft:Liberty` are rarely triggered, each accounting for less than 1%. Surprisingly, the value of `mft:Cheating`, despite its semantic proximity to `mft:Betrayal`, is consistently present.

This diagram should be compared with Figure 5.2, which displays the original annotation of the MFRC corpus. It is important to note that Figure 5.2 represents the full dataset of non-unique sentences, each of which was annotated by at least three annotators. Therefore, the occurrence count for each label far exceeds the aforementioned subset.

The initial observation is that in its original form, 45% of the dataset is labeled as not containing any MFT values. Additionally, a further 16% is categorized as "Thin Morality," which, as discussed in Chapter 3, indicates that some form of morality is detected but either (i) does not surpass a threshold set ad hoc by the annotator or (ii) represents a value not covered by the MFT dyads. It is worth noting that the MFRC dataset adopts the MFT terminology in its 2022 version, as explained in Section 3.2.4: `mft:Sanctity` is labeled as "purity," and `mft:Fairness` is further specified into two subclasses of fairness: "Equality" and "Proportionality."

The graph-based MFT annotation diagram presented in Figure 5.1 does not include the "Non-Moral" label. As explained in Section 3.3.3, the graph-based model does not have any non-moral triggers. Therefore, it can be inferred that if a certain theory's trigger is not retrieved in the automatic annotation, the corresponding value is not detected in the sentence.

In addition to offering a broader range of activations compared to the MFT framework, the graph-based detector also captures values from the BHV theory and the Folk module.

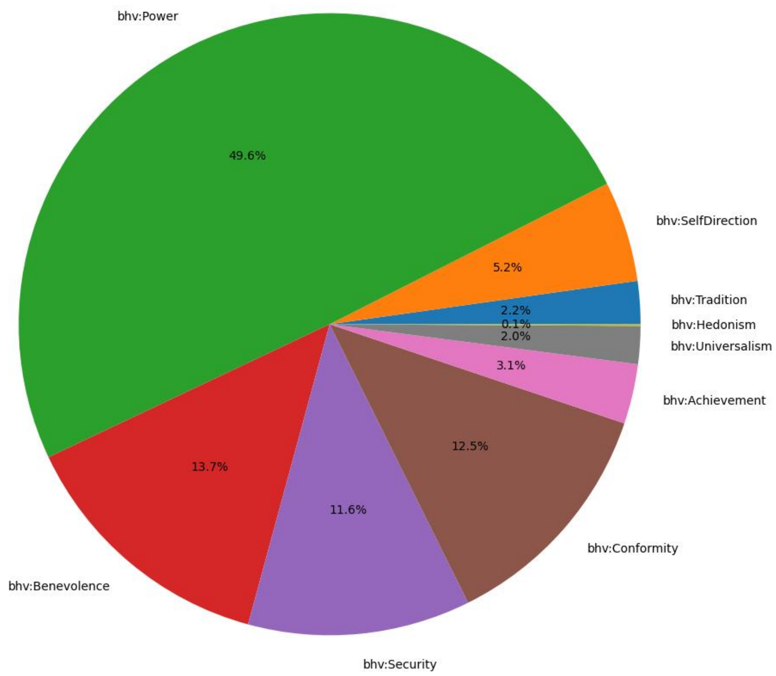


Figure 5.3: BHV values automatically detected in MFRC corpus.

The diagram in Figure 5.3 displays the BHV activations.

One notable observation is that 9 out of 10 values from the value wheel are detected, with the exception of `bhv:Stimulation`, which has limited lexicalization and thus evokes less frequently. The value of `bhv:Power` is the most frequently detected, likely due to the breadth of its trigger graph and its alignment with the topics covered in the MFRC dataset. In the subsequent sections, a detailed analysis of co-occurrences among values from different modules, emotions, and image schemas will be provided, along with proposed value clusters and thorough explanations.

Figure 5.4 presents a bar chart displaying the number of occurrences for Folk values, identifying `folk:Leadership` as the most frequently detected. This finding aligns with the other two diagrams and the semantics of the MFT and BHV modules, forming a cluster of values associated with social, institutional, and public leadership roles that correspond to the primary topics present in the MFRC dataset.

Figure 5.5 illustrates the percentage of activation for image schemas, providing partial confirmation of the hypotheses regarding the pervasiveness of certain image schemas such as Containment and Source_Path_Goal, which are frequently expressed. Given the numerous comments in the MFRC dataset addressing potential political developments

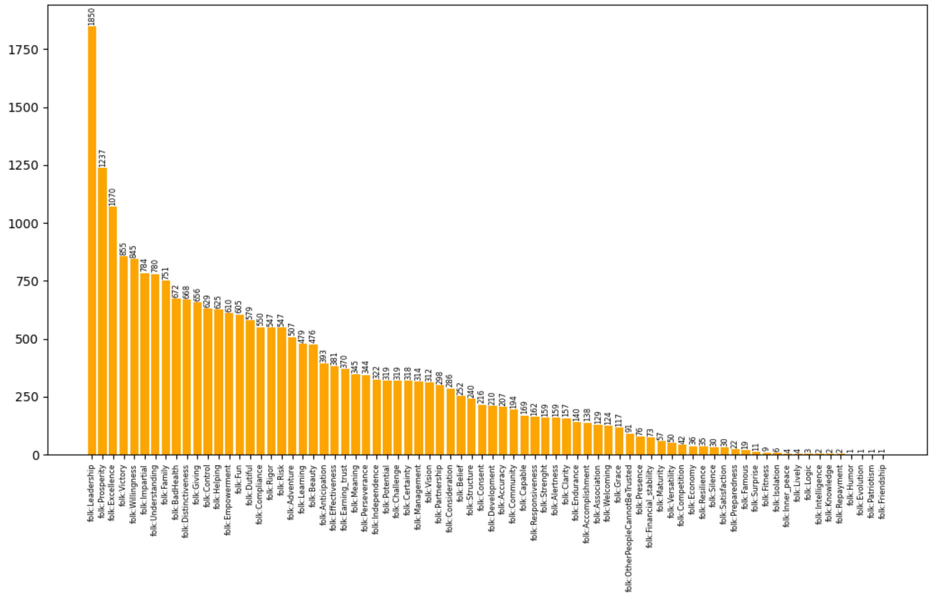


Figure 5.4: Folk values automatically detected in MFRC corpus.

and different global scenarios, the image schema On_Path_Toward exhibits a substantial activation with 2,482 occurrences.

Finally, Figure 5.6 displays the triggering of Basic Emotions, which, although fewer in number, is proportional to the size of the Basic Emotions graphs and the extent of semantic coverage discussed in Table 4.1 in Chapter 4.

These diagrams and tables serve to provide an overview of the detected semantics by the unified detector. The subsequent sections delve into the detailed interactions among these different levels of knowledge and demonstrate how the graph structure can be leveraged for new inferences and the extraction of novel knowledge.

5.2 Analysis and Discussion

In this section, we leverage the graph structure of the data generated from the MFRC experiment to uncover the following: (i) situations where image schemas, emotions, and values co-occur, (ii) the semantic meaning and structure of these co-occurrences, and (iii) potential patterns that can enhance inference capabilities. As demonstrated in the previous section, the true power of our methodology lies not in simple string labeling, but in the ability to reason over the graph structure.

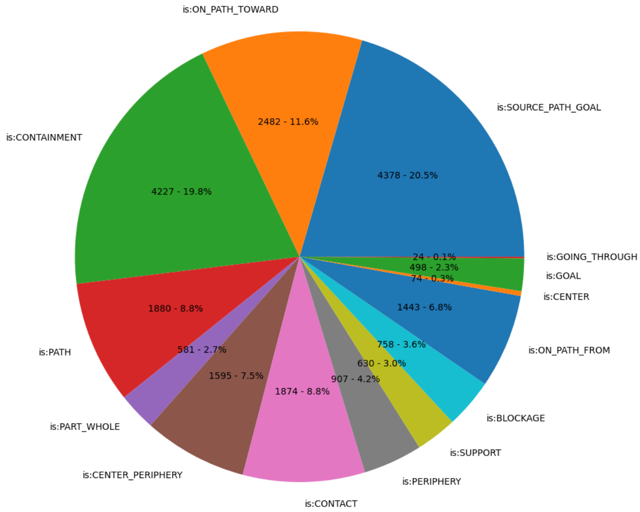


Figure 5.5: Image Schemas automatically detected in MFRC corpus.

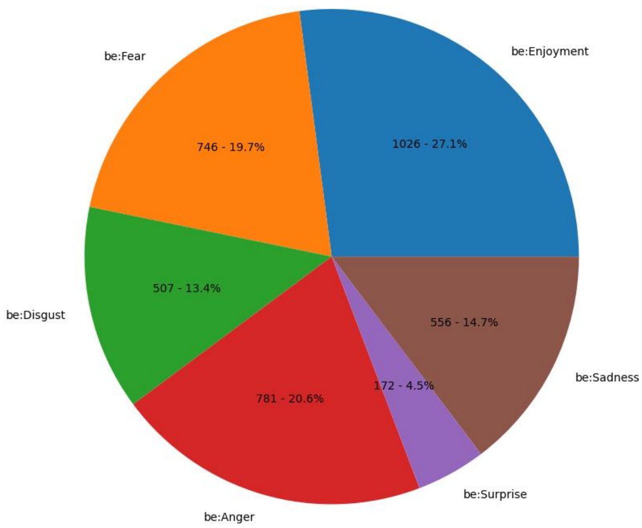


Figure 5.6: Basic Emotions automatically detected in MFRC corpus.

5.2.1 Schematic Integrations

This section examines image schematic situations, which refer to situations in the world that exhibit a semantic structure satisfying certain image schemas, where the image schemas take on semantic roles associated with value triggers. Mandler refers to constructs composed of non-image-schematic elements organized by an image schema structure as "Schematic Integrations," as discussed in Chapter 2, Section 2.2.3. According to Mandler, these integrations represent 'the first conceptual representations to include non-spatial elements, by projecting feelings or non-spatial perceptions to blends structured by image schemas' [215]. It should be noted that we are not asserting that all occurrences of a value trigger, qualified by or associated with an image schema, constitute a schematic integration. Each case must be considered individually. Nonetheless, this methodology lays the groundwork for future research directions, as outlined in the concluding section of this chapter.

Mandler specifically connects emotions to embodied cognition. However, as discussed in Section 3.1 based on Rozin et al. [277], emotions and morality cannot be neatly separated. The MetaNet repository reveals that over 20 conceptual metaphors are directly linked to the notion of "morality," such as `MORAL CORRUPTION IS A DESTRUCTIVE FORCE` or `MORALITY IS PURITY`, while several metaphors are associated with concepts grounded in values from one or more specific frameworks, such as `DEMOCRACY IS AN EQUAL PARTNERSHIP` or `TRUST RELATIONSHIPS ARE BUILDINGS`. Many of these conceptual metaphors related to morality are, in fact, image-schematic, such as `MORALITY IS UP`, `IMMORALITY IS DOWN`, and `MORALITY IS A STRAIGHT PATH`. In this section, we analyze each case individually, referring, when applicable, to the associated conceptual metaphor.

Two questions guide the investigation of the graphs:

- Q1: Is there any image schema situation where any value trigger is involved?
- Q2: If so, what is the relationship between the image schema and the value?

To provide an example of the commonsense knowledge explored in this manner (as well as in the subsequent sections) and to demonstrate how we can answer Q1 and Q2, let's consider the following sentence: 'The terrible killer was finally stopped by the police.'

Figure 5.7 illustrates the semantic dependencies structure. The root node represents the main verb of the sentence, "stop," depicted as an individual (violet square) belonging to the `fred:Stop` class (yellow circle). This class represents situations that satisfy the action of "stop" and is a subclass of various nodes sourced from semantic web resources: the `fsdata:Halt`, `fsdata:ProcessStop`, and `fsdata:ActivityStop` frames, the VerbNet `vn:Stop_55040100` verb, and the DOLCE upper layer ontology `dul:Event` class. The `fsdata:Halt` frame (and the `vn:Stop_55040100` verb, not shown here for clarity) activates the `is:BLOCKAGE` image schema. The original root node, indicated as an IS trigger

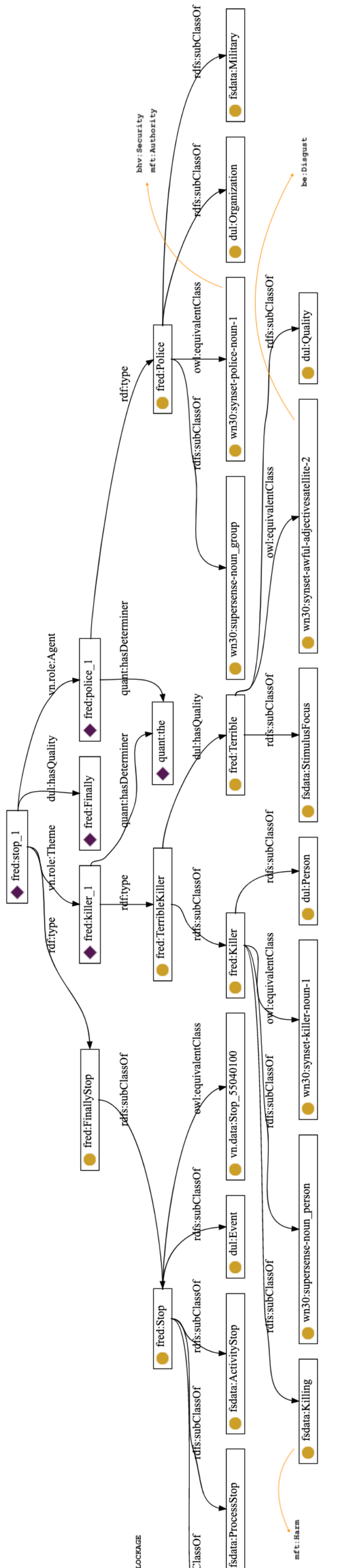


Figure 5.7: IS-emotion-value automatic detection for the sentence: "The terrible killer was finally stopped by the police."

through the aforementioned graph path, takes on the role of Theme with the “killer” node. The “killer” node, following the edges and nodes, evokes the `fsdata:Killing` frame, which in turn serves as a trigger for the `mft:Harm` value. Additionally, the killer is modified by an adjective disambiguated to the WordNet synset `wn:awful-adjectivesatellite-2`, which triggers the `be:Disgust` emotion according to Ekman’s model. Finally, the root node “stop” also functions as the Agent with the “police_1” node, disambiguated to the WordNet synset `wn:police-noun-1`, which triggers the `mft:Authority` and `bhv:Security` values. The IS profile of this sentence is composed of Blockage. The emotion profile is composed of `be:Disgust`. The value profile is composed of `mft:Authority`, `bhv:Security`, and `mft:Harm`. But how are these triggers related to each other? By leveraging the graph structure with our frame-based approach, the knowledge structure that emerges is as follows:

> Some trigger for `mft:Authority` and `bhv:Security` is Agent in a Blockage situation, whose Patient is a `mft:Harm` trigger, qualified emotionally by a `be:Disgust` trigger.

Furthermore, by leveraging previous ontological structures, we can assign roles to the `mft:Harm` trigger and the trigger for `mft:Authority` and `bhv:Security` within the Blockage situation. The `mft:Harm` trigger, acting as the Patient in the Blockage situation, assumes the role of `is:BLOCKED`, while the trigger for `mft:Authority` and `bhv:Security`, functioning as the Agent in the Blockage situation, assumes the role of `is:BLOCKER`.

Moreover, it is possible to extract a particular epistemic stance by utilizing commonsense knowledge that when something is qualified by a negative emotion, it likely carries a dysphoric semantic burden, at least from the speaker’s perspective. In this case, the qualification of the `mft:Harm` trigger as emotionally loaded with `be:Disgust` leads us to infer a negative epistemic stance towards the `mft:Harm` trigger, influenced by the presence of the `be:Disgust` moral emotion. Consequently, after identifying the roles `is:BLOCKER` and `is:BLOCKED` for the Blockage situation, it is plausible to assume a positive stance towards the `bhv:Security` and `mft:Authority` triggers for being the blockers of an entity associated with a negative epistemic stance.

It is important to note that this second assumption is highly conjectural at this stage and cannot be generalized to imply a positive opinion or perspective towards the `bhv:Security` or `mft:Authority` values per se, or their triggers. It simply indicates that, in this specific `vc:ValueSituation`, a negative epistemic stance is expressed towards a harmful event, allowing us to hypothesize a positive epistemic stance towards the `is:BLOCKER` of a dysphoric entity.

These initial paragraphs primarily focus on identifying patterns that involve image-schematic occurrences where value triggers participate, along with specific VerbNet roles, as exemplified above.

To retrieve image schematic situations where value triggers assume roles, the following SPARQL query is executed for each generated graph from each MFRC sentence:

```
SELECT DISTINCT ?class1 ?class2 ?value ?txt ?role
WHERE {
?g <https://w3id.org/sdg/meta#graphFor> ?txt .
?node1 rdf:type/owl:equivalentClass ?class1 .
?class1 isnet:activates isaac:BLOCKAGE .
?node1 ?role ?theme .
?theme (<>|!<>)* ?class2 .
?class2 vcvf:triggers ?value .
FILTER(regex(str(?role),
   'http://www.ontologydesignpatterns.org/ont/vn/abox/role/')) }
```

Line by line, in natural language, this query can be interpreted as follows: select all entities from a graph `?g` generated by some text `?txt`, where a certain `?node1` is an individual of a specific `?class1`. Additionally, the `?class1` node must be a subclass of an entity that activates an image-schema (in this query, it is Blockage, but it can be any image-schema). Furthermore, the `?node1` must have a VerbNet role relation that leads, through an unspecified number of nodes, to a `?value`. The number of nodes (N) is not specified since, at this level of the FRED graph, the remaining relations typically involve only type or subsumption relationships. This query reproduces the simplified version of the graph structure shown in Fig. 5.7. Upon examining Fig. 5.7, it can be observed that there are three nodes connecting the root node "stop_1" and Blockage. However, in this query, we consider only two nodes. This is due to the fact that the node "FinallyStop" in Fig. 5.7 is a fictitious node generated by FRED, which is then disambiguated one layer deeper. As the exact conditions for generating fictitious nodes are unknown due to internal FRED heuristics, we consider only the direct path from the root to the image-schema. This inevitably results in a loss of recall, but it ensures that the desired structure is matched and eliminates graphs that do not conform to this pattern. Next, we analyze all the retrieved matches for the specified pattern in the query. The query is repeated for each image-schema covered in the ImageSchemaNet module, querying all the graphs generated by the MFRC sentences.

Blockage The first case we consider is Blockage. According to the aforementioned SPARQL query, a co-occurrence of Blockage and value-related semantics in the MFRC data is retrieved in 86 graphs. The total number of occurrences, accounting for the possibility of multiple instances of the pattern within a graph, is 373. It is important to note that this number does not represent the total number of IS activations retrieved, which is shown in Figure **??**, but rather refers to a very specific pattern. The complete tables displaying the full list of graphs are available in the dedicated folder of the ValueNet GitHub repository.[1]. An interesting example is Graph 13.019: 'I know Macron was able to put out a statement 4 minutes before the law blocking election coverage went into effect. Maybe the story was posted before 00:00 last night as well.' In this sentence, the underlined part demonstrates the organization of knowledge according to image-schema. The Blockage situation, triggered by `vn:Block_67000000`, involves the WordNet synset `wn:law-noun-1` as the Agent, which serves as a trigger for `mft:Fairness` and `folk:Rigor`, and the synset `wn:election-noun-1` as the Theme. The underlying abstract structure that can be automatically retrieved from this example is as follows: `mft:Fairness` (in its "justice" or "law" extensional meaning) and `folk:Rigor` impose restrictions on the actions of some form of `mft:Authority`. This example is closely related to the cognitive metaphor `PREVENTING ENACTMENT OF LEGISLATION IS BLOCKING MOTION`.

Based on this specific example, we can hypothesize a general pattern: in a Blockage situation, the Agent, referred to in image-schematic terms as the `is:BLOCKER` (the entity that restricts movement), is negatively appraised and receives a dysphoric qualification from the Patient, represented by the `is:BLOCKED` entity.

Support There are 87 graphs matching this pattern for 374 occurrences. For instance, in Graph 5929 generated from the sentence 'I meant no Le Pen as president. But feel free to feed your persecution complex,' the trigger `wn:feed-verb-1` activates the Support value, with the Recipient role assigned to `wn:persecution-noun-1`, which triggers `mft:Harm`. This case is interesting because the supported entity carries a negative connotation due to its association with harm. Although "feed" is not a prototypical trigger for an image schema, it aligns with our semantics by operating at the synset level and is coherent with the underlying abstraction layer of the lexical representation. The Support trigger graph seems to overlap with triggers for `mft:Care`, `bhv:Benevolence`, and `folk:Helping` values. Hence, a general pattern that can be derived from this specific case is that in a Support situation, the Agent (referred to as the `is:SUPPORTER`) exhibits a positive epistemic stance towards the Patient (represented by the `is:SUPPORTED` entity).

[1] MFRC experiments material is available on the ValueNet GitHub: `https://github.com/StenDoipanni/ValueNet`

On_Path_From The On_Path_From image schema represents movement *away from* an entity with emphasis on the source. It is found in 270 graphs with 785 occurrences. For example, Graph 12.670 is generated from the sentence 'We The People she says, invoking the preamble while spewing dangerous misinformation. Ugh, this age of misinformation is getting much out of hand.' Although "spewing" is not prototypical, it does involve some form of movement. The VerbNet entity vb:Spew_43040000 triggers the On_Path_From image schema, along with the be:Disgust emotion and the mft:Degratation value. In this graph, the entity experiencing the *movement from* situation is represented by the Theme role wn:misinformation-noun-1, which triggers mft:Harm and is further qualified by wn:dangerous-adjective-1, activating folk:Risk. The knowledge structure of this graph identifies a metaphorical movement that carries a negative polarity, triggered by the be:Disgust emotion and the moral load of mft:Degratation. The entity involved in the movement triggers mft:Harm and is qualified by a folk:Risk trigger.

On_Path_Toward The On_Path_Toward image schema, in contrast to the previously mentioned image schema, places a greater emphasis on the Goal role. It is observed in 452 graphs with a total of 1,141 occurrences.

For example, in Graph 608 generated from the sentence 'Seems Trump wants France to leave the EU. Le Pen says she wants to leave the EU. Although Trump has done a U-Turn and is now for NATO. I don't buy the whole Putin - Le Pen love affair angle,' the relevant elements for this image-schematic situation are underlined. The trigger for On_Path_From is wn:leave-verb-1, with the Agent role assigned to the DBpedia entity db:France, which plays a role in the fs:PoliticalLocales frame. This frame triggers the mft:Authority value. Additionally, there is a Theme role represented by "EU," correctly disambiguated as db:EuropeanUnion, which again evokes the fs:PoliticalLocales frame and thus activates the mft:Authority value. Lastly, the trigger wn:love-noun-1 corresponds to the mft:Care value. This example highlights a departing movement between authorities, where entities possessing socio-culturally recognized importance move away from each other. It exemplifies the cognitive metaphor INSTITUTION IS A VEHICLE or INSTITUTIONS ARE PEOPLE, with one authority being the source and the other authority being the moving object.

Going_Through On the other hand, the Going_Through image schema only matches the desired pattern in Graph 5076, occurring twice. It is generated from the sentence 'Not really. Obama was asking the teacher to excuse the kid. Macron is saying the kids are authorized to skip school. Not the same thing at all.' Without context, it is challenging to determine the exact meaning of this sentence. However, the verb "skip" is disambiguated as the `wn:jump-verb-13` WordNet synset, which corresponds to a Going_Through situation. Although the activation of this image schema is debatable, it is interesting that the Theme of this metaphorical movement is `wn:school-noun-1`, triggering the `mft:Authority` value. This example is noteworthy because it likely employs metaphorical language, where the "children" are not actual children, and the "school" is used to refer to some form of authority. This allows for reflection on the generalization of this specific case, where a metaphorical expression referring to movement, such as "getting away with it," indicates escaping potential punishment or evading the attention of an authority figure.

Containment The Containment image schema exhibits numerous occurrences that match the desired pattern, with a total of 1,324 occurrences spread across 461 graphs. However, there is some noise present due to questionable triggers. In Graph 9643, an interesting case is presented: 'You evil hate filled man! Hey, I may be evil and hate-filled, but I'm not... wait, what was the last one you said?' The trigger for Containment is `vb:Fill_45040000`, which could be considered an example of the conceptual metaphor THE BODY IS A CONTAINER FOR EMOTIONS. However, due to the passive construction of the sentence, the Patient of the filling is `wn:hate-noun-1`, whereas it should be the Theme, and the "man" should be the Container. Additionally, `wn:hate-noun-1` triggers `be:Anger` and `mft:Harm`.

Source_Path_Goal As for the Source_Path_Goal image schema, there are 497 graphs and 1,473 occurrences that match the desired pattern. In Graph 6720, derived from the sentence 'YTA [You're The Asshole] and you know it very well. You need to disregard the backwards laws in Saudi Arabia or so and do the right thing, which is to share with your sister equally. Otherwise, you can never call yourself a man ever again. And please, do not tell us you are not following the teachings of Islam when you are planning to profit from an Islamic law that treats women as half their brother.' The trigger for Source_Path_Goal is `vb:Follow_51060000`, and the individuated Theme roles include `wn:Muslim-adjective-1`, which triggers `mft:Sanctity`, and `wn:law-noun-1`, which triggers `folk:Rigor` and `mft:Fairness`. Additionally, `wn:brother-noun-1` triggers `folk:Family`. This example could be seen as an instantiation of the conceptual metaphor LAWS ARE LIVING ENTITIES.

Contact For the Contact image schema, there are 461 graphs that match the pattern, with a total of 1,373 occurrences. In Graph 3712, derived from the sentence 'Macron talks a lot but in a pocket of Putin now, deals are going to be signed, just like Turkey. EU democracy charade for the sake of love for socialism, and they know. All that play will backfire into close border, more despotic measures to protect socialism and kiss up with Russia. EU manipulates foreign policy of US through social media. Autonomous is a joke; it is all Russia.' Although this sentence contains multiple value triggers, the one that matches the pattern for a Contact situation is `wn:manipulate-verb-1`, which triggers Contact. Its Theme role is `wn:policy-noun-1`, which in turn triggers `folk:Rigor` and `mft:Fairness`.

On the other hand, no matches are retrieved for the Link image schema.

The differences in the number of matches can occur due to several reasons:

- The number of verbal triggers in the IS knowledge graph: The pattern investigated assumes that certain ISs are associated with specific verbs. The success of retrieving matches depends on the availability of verbs that ground their meaning in the investigated IS;
- The correct retrieval of roles by FRED: The accurate identification and extraction of roles by the FRED system influence the matching process;
- The domain being investigated: The MFRC dataset primarily consists of sentences related to the political domain. The preference for certain ISs over others can be observed from the conceptual metaphors repository, which indicates that different domains may exhibit variations in the usage of ISs.

5.2.2 Embodied Morality

Another aspect of schematic integrations that warrants further investigation is the co-location of image-schematic and value knowledge triggered by the same entity. This implies that, in the conceptualization of a particular concept, both embodiment and moral knowledge occur simultaneously. Within the MFRC corpus, numerous occurrences of co-location between image-schematic and value triggers are observed. The following paragraphs highlight some of the most notable clusters in this regard. It is important to note that the clusters presented here represent overlaps identified within the MFRC dataset, and their realization and experimental validation can be explored using the provided queries and code available on the GitHub repository[2]

Each paragraph describes a specific instance of co-location wherein an image schema and a value are triggered by the same group of entities. These groups may range from extensive collections of WordNet synsets or DBpedia entries to single verbs or entities. Both cases are of interest for different reasons: (i) large groups of entities help identify distinct

[2]The python code and queries to retrieve image-schematic, emotion and value activators is freely available on the ValueNet repository here: `https://github.com/StenDoipanni/ValueNet/tree/main/ValueNet_code`

semantic domains and can be further explored through SPARQL querying of the graphs in which these triggers occur, and (ii) single entities that co-activate image schemas, values, and emotions point towards a direction akin to the Moral Molecules concept discussed in the MM module in Section 3.2.7. Such cases indicate that the semantics of complex concepts or situations can be understood by examining the simultaneous activation of multiple image-schematic, value, and emotion frames. Finally, if any cognitive metaphor is found to be grounded in the identified multi-layer knowledge overlaps through manual analysis, it will be provided in the subsequent paragraphs.

Support, Benevolence, Helping, and Care The concept of support, involving the expenditure of energy to physically maintain the position of someone or something, exhibits overlapping elements in both the image-schematic and value domains. Several triggers commonly associated with the semantics of *support* align with those related to *help*, such as `wn:help-verb-1`, `wn:aid-noun-2`, `wn:endorsement-noun-5`, `wn:feed-verb-1`, `wn:backing-noun-1`, the `fs:Assistance` frame, `wn:helpful-adjective-1`, `wn:provide-verb-2`, and `nurse-noun-1`. This co-location of image-schematic and value knowledge indicates an overlap in semantic role structure. The Theme/Patient of a benevolence, care, or `folk:Helping` situation also assumes the role of the `is:SUPPORTED` entity in an image-schematic support situation, while the Agent fulfills the `is:SUPPORTER` role. This structure, automatically extracted through SPARQL querying of the MFRC graphs, corresponds to the cognitive metaphors `ASSISTANCE IS SUPPORT` and `HELP IS PHYSICAL SUPPORT`, where the concepts of *Assistance* and *Help* are encompassed by the value frames of care, benevolence, and `folk:Helping`.

Interestingly, another semantic aspect is expressed by triggers such as `wn:blessing-noun-1` and `wn:gratify-verb-2`. These triggers offer additional interpretive insights into this semantic overlap, highlighting the provision of lasting care that can be recalled and utilized when needed. These triggers qualify an entity as being supportive. One example that exemplifies this overlap is seen in Graph 357, which is generated from the sentence: 'Yeah, I think most European leaders understood that. If you remember how Merkel/Macron spoke very harshly about Trump recently and <u>got big approval for it</u>.'

Support and Perseverance This cluster revolves around the entity `wn:sustainable-adjective-1`, which represents the conceptualization of a sustainable situation. An example illustrating its usage can be found in Graph 4238, which features the sentence: 'Also, a huge welfare state is not necessarily a <u>sustainable situation</u> under globalization. I'll sooner vote for Melenchon/Le Pen and burn the EU to the ground than give up the welfare state.'

Blockage, Power, Rigor, Fairness, and Authority This cluster encompasses various conceptualizations of limitations imposed by forms of superior control or authority. Triggers within this cluster include vb:Ban_67000000, wn:prohibition-noun-2, wn:prevent-noun-2, vb:Forbid_67000000, and vb:Outlaw_29070000. These triggers contribute to the realization of the cognitive metaphor CONTROL IS PHYSICAL RESTRAINT.

A closely related cluster is the Blockage and bhv:Security overlap, which includes the trigger wn:imprison-verb-1. Additionally, there is a slight variation in meaning with the Blockage and folk:Consent overlap. The triggers wn:limit-noun-1 and wn:restrictive-adjective-1 activate both the Blockage and folk:Consent frames. The notion of consent represents a more recent addition to the folk value list, encompassing the idea of having control over one's body, freedom of choice, faith, and behavior. It is considered an oblique value and is classified as a subclass of bhv:SelfDirection in the ValueNet ontology.

Blockage, Harm and Oppression This cluster encompasses entities that represent harmful and impeding movement situations. The presence of wn:crisis-noun-1 indicates an overlap between Blockage and mft:Harm, while triggers such as wn:jail-noun-1, wn:detention-noun-1, wn:prison-noun-1, and the fs:Prison frame illustrate the overlap between Blockage and mft:Oppression. This cluster reflects the negative form of the cognitive metaphor FREEDOM OF ACTION IS THE LACK OF IMPEDIMENTS TO MOVEMENT.

Contact and Power In the Contact and Power overlap, the trigger wn:manipulate-verb-1 represents the idea of being in contact and exerting control over actions or events. This realization corresponds to the cognitive metaphor CONTROL IS OBJECT MANIPULATION.

Contact and Fun Another overlap is observed between Contact and Fun, which involves the folk:Fun value. This overlap is represented by two entities, namely snog-verb-1 and vb:Kiss_36020000. These triggers highlight a sensory pleasure dimension and point to common knowledge associated with more conservative and restrictive cultural contexts. In these scenarios, Contact between individuals may be forbidden or discouraged to prevent "immoral" sensory pleasure. The fact that these triggers also activate the be:Enjoyment frame further supports this interpretation.

Contact, Loyalty, and Presence The folk:Presence value is centered around the idea of being there for someone, emphasizing emotional closeness. This small cluster consists of the triggers vb:Tie_22040000 and wn:connect-verb-1, representing the cognitive metaphor INTERRELATEDNESS IS PHYSICAL INTERCONNECTEDNESS.

Contact and Harm In the Contact and Harm cluster, common knowledge related to acts of violence involving physical contact and harm is included. Triggers such as vb:Wallop_18030000, vb:Punch_18020000, wn:slam-verb-2, wn:knock-verb-1, and wn:beater-noun-2 are part of this cluster.

Containment and Experience The Containment and Experience overlap includes the notion of wn:experience-noun-1 triggering both Containment and folk:Experience. This can be explained by the cognitive metaphors EXPERIENCING A STATE IS POSSESSING AN OBJECT and MIND IS A CONTAINER.

Containment and Care The Containment and Care cluster consists of two entities with different conceptualizations grounded in image-schematic and moral knowledge. The trigger wn:safe-noun-1 represents keeping goods secure and hidden in a container, while wn:condom-noun-1 represents a more progressive perspective, highlighting its morally loaded semantic nature that can be seen as blameworthy or praiseworthy depending on different faction

On_Path_From and Degradation In this cluster, one of the triggers is vb:Spew_43040000, which was mentioned in the previous section and also triggers be:Disgust. Another interesting entity is vb:Die_48020000, which represents the act of departing from this life and carries a heavy moral-emotional connotation related to the sanctity-degradation moral foundations theory (MFT) dyad and the emotion of be:Disgust as defined by Ekman.

Source_Path_Goal and Control The Source_Path_Goal and Control cluster consists of elements that involve movement and the purpose of that movement. The triggers in this cluster are wn:nudge-verb-1, wn:jostle-verb-2, wn:slap-verb-1, and wn:yank-verb-1.

Source_Path_Goal and Harm In the Source_Path_Goal and Harm cluster, the entities represent harmful intentions that involve movement. An example is wn:invade-verb-1, which is often used in far-right communication to connote phenomena such as migration. Interestingly, this synset is emotionally loaded and also triggers the emotion of be:Anger.

5.2.3 Moral Emotions

Specular to the IS situation detection, we slightly modify the aforementioned SPARQL query to retrieve emotion situations where the VerbNet role is triggered by a value. The query is as follows:

```
SELECT DISTINCT ?class1 ?class2 ?value ?txt ?role
WHERE {
?g <https://w3id.org/sdg/meta#graphFor> ?txt .
?node1 rdf:type/owl:equivalentClass ?class1 .
?class1 emo:triggers be:Anger .
?node1 ?role ?theme .
?theme (<>|!<>)* ?class2 .
?class2 vcvf:triggers ?value .
FILTER(regex(str(?role),
    ''http://www.ontologydesignpatterns.org/ont/vn/abox/role/'')) }
```

In the example query, we retrieve occurrences of the `be:Anger` frame, but it can be replaced with any of the basic emotions. Exploring the co-location of other types of knowledge and emotions allows us to draw conclusions about epistemic stance and polarity attribution when possible.

Anger Running the query on the MFRC graphs, we find that the anger frame occurrence, with a role triggered by a value, is matched in 138 graphs, totaling 386 occurrences. Here, we provide a meaningful example from Graph 12,896, generated from the following string: "That's exactly the point. We hate making other people rich and being forced to work in order to survive or live comfortably".

Fig. 5.8 illustrates a reduced version of the graph generated by the IS, emotions, and value detector. Due to visualization constraints, some alignments and disambiguations are omitted for clarity.

In the shown figure, the root node "hate_1" leads to `vb:Hate_31020100`, which triggers the moral emotion `be:Anger` and the `mft:Harm` value. This matches the pattern searched with the query, as the root node takes the VerbNet Theme role of "situation_2", which, in turn, triggers some value in its semantic themes role. The term "force" is disambiguated as the VerbNet verb `vb:Force_59000000`, which is a trigger for `mft:Oppression`. The node "survive" is disambiguated as `vb:Survive_47010110`, which triggers `folk:Endurance`. Lastly, the lexical unit "rich" leads to the triggering of `folk:FinancialStability`. Although `folk:FinancialStability` should arguably

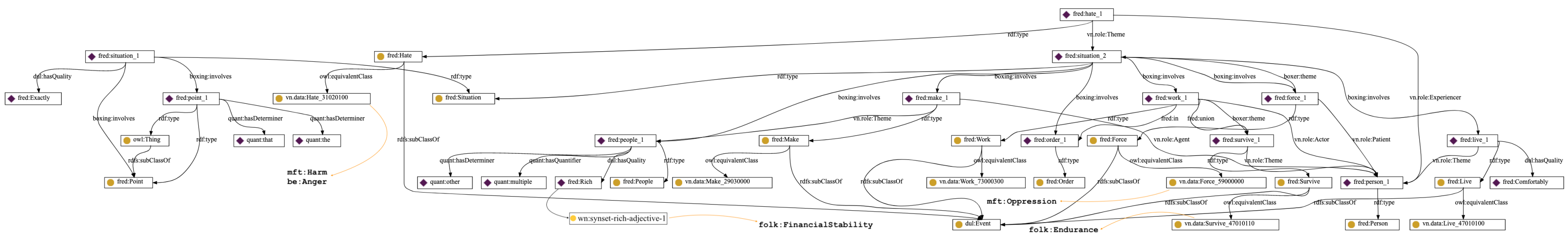


Figure 5.8: IS-emotion-value automatic detection for the sentence: "That's exactly the point. We hate making other people rich and being forced to work in order to survive or live comfortably."

not set a moral compass, it is considered a value in the Folk module, representing a daily-life goal and shaping individual behavior.

The abstraction of the moral structure for this sentence is as follows: a "person_1" is the Experiencer of `be:Anger` and `mft:Harm`, with three Theme roles: `mft:Oppression`, `folk:Endurance`, and `folk:FinancialStability`. From the analysis of the value schema, we could hypothesize an anti-capitalistic position, which aligns with the actual meaning of the sentence that generated the graph.

Expanding on this, based on the formal semantics and considering that `be:Anger` is a negative emotion and `mft:Harm` is the negative pole of the MFT dyad, we can infer a negative epistemic stance for the expressor towards the themes of `mft:Oppression`, `folk:FinancialStability`, and `folk:Endurance`. In particular, `folk:FinancialStability` aligns with the BHV theory, being declared a subclass of `bhv:Achievement`.

Referring to the BHV Value Wheel (Fig. 3.2) in Sec. 3.2.2, we can further infer that a negative stance towards the negative pole of an MFT dyad and towards a value in a certain quadrant of the BHV value wheel circumplex model implies a positive stance towards their opposites, specifically `mft:Liberty` and `bhv:Universalism`. Taking a broader perspective, we could infer that in this specific circumstance, "person_1" has a positive epistemic stance towards `bhv:SocialFocus` and `bhv:GrowthAndAnxietyFree` attitudes. The general heuristic derived from this case is as follows: if an entity is the Experiencer of a negative emotion and the Theme is a value, then that value is marked as dysphoric, and the Experiencer is considered to have a negative epistemic stance towards the value (specifically towards the node that triggers it).

Disgust A total of 238 occurrences across 87 graphs were found, matching the desired pattern for disgust situations. An intriguing case for analysis is Graph 16, which originates from the following sentence: 'Valls is such a disgusting traitor to his own party. He helped Macron to win, thought.' The underlined phrase in the sentence indicates the presence of a disgust situation. The adjective "disgusting" is disambiguated as `vb:Disgust_31010000`, while its Cause VerbNet role is assigned to a node that is disambiguated as `wn:traitor-noun-1`, triggering the concept of betrayal. Thus, the moral content structure of the sentence implies a negative epistemic stance towards betrayal, inferred from its association with disgust.

It is noteworthy that Annotator00 and Annotator03 labeled this case with *Loyalty*, while Annotator04 associated it with *Sanctity*, *Loyalty*, and *Authority*. Our detector identified the presence of the disgust emotion, as well as the values of benevolence, betrayal, care, `folk:Helping`, and `folk:Victory`.

This case represents an emblematic example where loyalty is not explicitly or factually mentioned in the sentence, but the expressor conveys a form of negative evaluation towards its opposite. By employing our method, we are able to assign significance to the given labels and deduce the presence of the loyalty value through the co-location and semantic dependency between disgust and betrayal. Consequently, the general heuristic derived from this sentence is as follows: if a value serves as the cause of a negative emotion, it is deemed dysphoric, potentially leading to a positive epistemic stance towards its opposite.

Enjoyment A total of 508 occurrences across 170 graphs were found to match the Enjoyment pattern. For the purpose of analysis, we examine Graph 5975, which is derived from the following sentence: 'No, the Le Pen/Putin gang only know a shallow nationalism that thrives on hatred and cannot be but destructive. Macron is the true patriot because he is prepared to give his best for the country.' In this graph, an enjoyment situation is triggered by `vb:Thrive_39060000`, evoking `folk:Prosperity`. The Patient role is assigned to the word "hate," disambiguated as `wn:hate-noun-1`, which triggers both `mft:Harm` and `be:Anger`. Unfortunately, in this particular case, the term "nationalism" is not disambiguated to any specific resource, resulting in a generic local individual `:nationalism_1` being assigned as the Agent for the enjoyment situation. Consequently, no particular epistemic stance can be extracted from this example. The only noteworthy aspect is the retrieval of a `folk:Prosperity` situation with a negative value and emotion as its Theme. If a Patient, Undergoer, or Experiencer had been expressed, it might have allowed for the detection of a more complex emotion such as `occ:Gloating`, involving taking pleasure in negative events involving others rather than oneself. However, this is not the case here.

Fear Moving on to Fear situations, 83 occurrences across 36 graphs satisfied the desired pattern. Graph 2917 presents an interesting case study derived from the sentence: 'The EU is scared. They cannot afford a Le Pen victory because they would lose France. This, coupled with the loss of the UK, would kill the EU.' The Fear situation is triggered by `vb:Scare_31010000`, with its Experiencer role filled by the "EU," disambiguated as the DBpedia entity `db:EuropeanUnion`. It evokes the `fs:PoliticalLocales` frame, which serves as a trigger for `mft:Authority`. The provided example illustrates the realization of the cognitive metaphors `GOVERNMENT IS A PERSON` and `POLITICAL PARTIES ARE PEOPLE`, where the domains of *government* and *political parties* are generalized to encompass any form of institution triggering the concept of authority.

Sadness The Sadness frame, as specified in the query, is found in 188 occurrences across 56 graphs within the MFRC dataset. Graph 10.703 provides an illustrative example derived from the sentence: 'Another outcome for decades of being shamed for your heritage.' This case offers a straightforward scenario that allows for further inferences. The Sadness situation is triggered by `Shame_33000000`, with the `wn:heritage-noun-1` WordNet synset serving as the Predicate role, which triggers the concept of `bhv:Tradition`.

This case is particularly relevant since literature suggests that *shame* is one of the primary moral emotions, heavily influenced by socio-cultural variables. In fact, a Sadness situation can be seen as the result of a preceding situation. While Ekman's Basic Emotions tend to flatten many complex emotions into the six basic ones, a more detailed analysis of Sadness, considering it as the cognitive state of the Patient resulting from a shame situation, can make the latent moral-emotional content of the dynamics more explicit. Furthermore, a shame situation associated with the `bhv:Tradition` value may indicate the presence of *prejudice*, aligning with the moral molecules approach proposed by Curry [42].
Lastly, no matches were retrieved for the Surprise emotion within the specified query.

5.2.4 Emotions and Values Coactivation

We proceed with the analysis by presenting clusters of emotions and values coactivated by the same entities, aiming to identify clusters of meaning associated with moral emotions. It is important to note that these clusters represent only the occurrences matching the specified pattern in experiments conducted on the MFRC dataset, and they do not encompass all the overlaps between graphs. Many entities evoke different emotion and value frames, even though their semantics may differ—for example, they can indicate a physiological manifestation of an emotion or the absence of a particular value. However, despite these differences, it is relevant to highlight these overlaps as they constitute the semantic connection between frames in different knowledge layers. While multiple examples are typically retrieved, we provide the ID of three graphs after each cluster description for further exploration of the data.

Surprise and Anticipation The `folk:Anticipation` value revolves around the concept of estimating the plausibility of an event to be prepared for it. Both Surprise and Anticipation share a common trigger, which is `wn:unexpected-adjective-1`. (e.g., Graphs 5354, 12.300, 309)

Surprise and Clarity The essence of the `folk:Clarity` value lies in being clear, comprehensible, and devoid of hidden pitfalls or ulterior motives. The intriguing cluster connecting Surprise and `folk:Clarity` is associated with
`wn:confusing-adjectivesatellite-2`. The semantic overlap between the folk value

and the emotion appears particularly fitting for this synset, as *being confused* can be conceptualized as perceiving something as unknown or not understanding it despite it should be comprehensible or clear. (e.g., Graphs 8961, 949, 11.172)

Surprise, Fun and Excellence The `folk:Excellence` value embodies the idea of excelling, surpassing mere satisfaction, and standing out from the average. These entities converge in `wn:amazing-adjectivesatellite-1` and `vb:Amaze_31010000`, which combine the element of novelty with the enthusiasm of positive appraisal represented by `folk:Fun`. (e.g., Graphs 7728, 9064, 9796)

Surprise and Surprise An intriguing observation is the pairing of `be:Surprise` as a basic emotion with the `folk:Surprise` folk value, which revolves around appreciating being surprised in life. The entities triggering both are `wn:surprised-adjective-1`, `wn:aghast-adjectivesatellite-1`, and `wn:amazed-adjectivesatellite-1`. (e.g., Graphs 772, 7436, 8770)

Anger, Power, and Control The interplay between the `be:Anger` basic emotion, the `bhv:Power` value, and the `folk:Control` value is particularly noteworthy in the case of `vb:Frustrate_31010000`. The experience of anger, in fact, arises from a perceived lack of power. It is important to note that the absence or lack of a value does not necessarily imply a bias towards its opposite pole (in the case of the MFT dyadic system) or quadrant (in the case of the BHV circumplex model). The cognitive state of *frustration* can be axiomatized in first-order logic as follows:

$$\texttt{frustrate(x)} \rightarrow \texttt{angry(x)} \quad \wedge \quad \neg\texttt{inPower(x)} \quad \wedge \quad \neg\texttt{inControl(x)} \tag{5.1}$$

This finding is significant as it introduces, from a bottom-up perspective, the idea of a *lack of value* as a determinant for a particular concept. This idea is not explicitly stated in the aforementioned theories. The general pattern that can be derived is that the absence of power or control can be the cause of negative emotions. (e.g., Graphs 5716, 6893, 12.152)

Anger and Fairness The `be:Anger` emotion and the `mft:Fairness` value converge in the concept of *revenge*, which is represented by `wn:retaliation-noun-1`, the `fs:Revenge` frame, and `wn:revenge-verb-1`. This finding aligns with the Curry Moral Molecule framework, where revenge is modeled as a subclass of the `mm:Reciprocity` social cooperation strategy. Revenge carries the emotional weight of anger and encompasses the notions of "an eye for an eye," "private justice," and "rebalancing the wrong suffered." (e.g., Graphs 9264, 3953, 495)

Anger and Harm Several triggers activate both the `be:Anger` emotion and the concept of `mft:Harm`, particularly in relation to the notion of *conflict*. These triggers include `wn:hate-verb-1` (also in its noun form), `wn:discord-noun-2`, `wn:contend-verb-6`, the `fs:Quarreling` frame, `wn:aggressive-adjective-1`, `wn:hostile-adjective-1`, `wn:fight-noun-2`, `vb:Argue_37060000`, and `vb:Retaliate_71000000`. (e.g., Graphs 3003, 3941, 55)

Disgust and Willingness The core concept of the `folk:Willingness` value is to be prepared and willing to undertake something. Triggers such as `wn:loath-adjectivesatellite-1` and `wn:reluctance-noun-2` are associated with both `be:Disgust` and `folk:Willingness`. (e.g., Graphs 4426, 12.571)

Disgust and Cheating The DBpedia entry `db:Bigotry` serves as a trigger for both the `be:Disgust` emotion and the concept of `mft:Cheating`, understood as *unfair treatment towards someone*. (e.g., Graphs 3272, 244, 740)

Disgust and Degradation This cluster is not only grounded theoretically, drawing on the psychology of disgust, but also supported by the data. Many triggers are shared by `be:Disgust` and `mft:Degratation`, including the previously mentioned `Spew_43040000`, `vb:Die_48020000`, `wn:fester-noun-1`, `wn:vomit-verb-1`, `vb:Decay_47020000`, the `fs:Rotting` frame, and even less directly related triggers like `wn:sin-noun-1` and the DBpedia entity `db:Disgust` itself. This cluster embodies the cognitive metaphor `COGNITIVE AVERSION IS PHYSICAL DISGUST`. (e.g., Graphs 7187, 9513, 998)

Disgust and Harm Triggers for both the `be:Disgust` and `mft:Harm` frames include `vb:Loathe_31020000`, `vb:dehumanize-verb-1`, and `wn:hater-noun-1`. The investigation highlights the notion of *hate* as a complex concept involving anger, disgust, and harm. (e.g., Graphs 1150, 383, 6457)

Enjoyment, Benevolence and Care Triggers that combine the concept of happiness with kindness, a caring attitude, charity, and mercy include `vb:Ease_45040000`, `vb:Reassure_31010000`, `wn:fulfillment-noun-1`, `wn:hug-noun-1`, and `wn:facilitate-verb-1`. (e.g., Graphs 3699, 5662, 772)

Enjoyment and Fun These two frames share semantic similarity, as their extensionality is satisfied by several overlapping entities, including wn:enjoy-verb-1, wn:pleasure-noun-1, wn:amusing-adjectivesatellite-2, vb:Satisfy_31010000, vb:Laugh_40020000, wn:happy-adjective-1, wn:cuddle-verb-1, and various entities related to physical manifestations of affection, externalization of joy, enthusiasm, and pleasure. (e.g., Graphs 7842, 326, 5039)

Enjoyment and Prosperity The concept of folk:Prosperity, which entails evaluating and appreciating a prosperous state, is closely related to be:Enjoyment. Lexical triggers for this cluster include wn:wellbeing-noun-1 and Thrive_39060000. (e.g., Graphs 5971, 2910, 6730)

Enjoyment and Harm Despite their apparent differences, these two frames are linked by entities such as wn:boast-verb-1, vb:Brag_37080000, and vb:Provoke_31010000. (e.g., Graphs 10.208, 1716, 3626)

Enjoyment and Liberty Entities such as wn:peace-noun-1 and wn:freedom-noun-1 contribute to the overlap between the frames of be:Enjoyment and folk:Liberty. (e.g., Graphs 1011, 7706, 1207)

Sadness and Endurance

The intersection of a negative emotional state with the value of folk:Endurance, which involves positively evaluating the ability to resist and persist, is marked by the lexical trigger vb:Suffer_31020000, denoting a state of prolonged suffering. (e.g., Graphs 925, 8177, 4607)

Sadness and Understanding

The folk:Understanding value is characterized by the idea of comprehending the suffering, contempt, and negative state of others without emotional commitment. The overlap between be:Sadness and folk:Understanding is represented by wn:commiseration-noun-1. (e.g., Graphs 5789, 10.655, 7867)

Sadness and Authority The presence of wn:shame-noun-2 in the overlap between be:Sadness and mft:Authority validates the theoretical framework of moral emotions. Shame, as a negative emotion, is closely tied to the social and cultural context in which it occurs. (e.g., Graphs 692, 1919, 11.512)

Sadness and Harm This cluster is particularly extensive, encompassing notions related to physical pain or suffering, including triggers such as wn:suffer-verb-2, wn:pain-noun-1, wn:agony-noun-2, and vb:Cry_40020000. As mentioned earlier, these triggers evoke both be:Sadness and mft:Harm frames in different ways. For example, wn:suffer-verb-2 is a clear trigger for be:Sadness, as it implies the occurrence of a previous harmful event that leads to the experience of sadness. (e.g., Graphs 4607, 11.161, 680)

Sadness and Subversion The combination of vb:Protest_71000000 and wn:protest-verb-1 connects the feeling of dissatisfaction, which causes sadness, with the active expression of subversion. (e.g., Graphs 2392, 10.315, 603)

Fear and Anticipation The ability to anticipate future events and the fear of negative outcomes are semantically intertwined in the lexical trigger wn:apprehension-noun-1. (e.g., Graphs 12.279, 11.021)

Fear and Risk The folk:Risk value revolves around appreciating the thrill of taking risks. Individuals who embrace this value tend to evaluate circumstances that evoke a sense of (mild or moderate) fear, such as wn:gamble-verb-1, the fs:RiskySituation frame, and wn:hazard-noun-1. (e.g., Graphs 4145, 7766, 8311)

Fear and Harm This cluster encompasses elements that pose threats to individuals or society, including triggers such as vb:Threaten_31010000, wn:panic-noun-1, db:Terrorism, wn:menace-noun-1, and more. (e.g., Graphs 8779, 6975, 2278)

In this investigation, we focused on exploring individual emotions and the co-occurrence of multiple values. However, the query can be easily modified to examine elements that activate multiple emotions simultaneously. For example, by emphasizing the euphoric-dysphoric dimension, it would be possible to explore the co-occurrence of specific clusters of values with general negative emotions like be:Sadness, be:Fear, be:Disgust, or be:Anger.

5.2.5 Value Scenarios

Section 5.2.2 explored the combination of IS and values, while Section 5.2.4 focused on emotions and values. In this section, we shift our attention to the interaction and influence of values on each other, the combination of their meanings in complex constructs, and the utilization of the graph structure to uncover new knowledge. The fundamental question addressed in this section is: What are the primary value relationships concerning other values, and how do they interact with each other?

To investigate this aspect, we conducted a similar query to the one presented in Section 5.2.2, but with a focus on value situations where other values play a role. Due to the extensive number of values and the lack of theoretical authority of folk values compared to MFT and BHV values, our query was refined to consider only MFT and BHV values as the main situations and all BHV, MFT, and Folk values as roles.

The main findings are outlined in the subsequent paragraphs.

Harm As discussed in Section 5.1.1, `mft:Harm` is one of the most frequently retrieved values due to its broad semantics. Specifically, within the MFRC, there are 2787 instances of harm situations spanning across 889 graphs. To provide an illustrative example, consider Graph 2834 generated from the following statement: 'Exactly my thoughts. I can understand Melenchon running and <u>hurting Macron's chances</u>, after all, they have enough differences, but Valls? <u>Valls will only hurt Macron</u>.' In this graph, two `mft:Harm` triggers are identified: `wn:ache-verb-3` and `vb:Hurt_40081000`. Both triggers assume the Experiencer role associated with the `fs:Leadership` frame, which serves as a trigger for `mft:Authority`. Thus, the value structure for this sentence entails a harm situation with an Experiencer represented by a manifestation of authority, specifically reflecting the cognitive metaphor `NATIONAL POLITICS IS A BATTLEFIELD`.

Care There are 1540 occurrences of `mft:Care` situations found across 462 graphs. We present two examples of `mft:Care` situations as they exhibit interesting patterns.

The first example is taken from Graph 7914, derived from the sentence: I would <u>love to artfully retort</u>, however, you are illiterate and cannot read this.' In this sentence, a `mft:Care` situation is evoked by the trigger `wn:love-verb-1`, with the `wn:retort-verb-1` and `fs:CommunicationResponse` frame serving as the Theme roles. Both of these entities also act as triggers for the `folk:Responsiveness` value. The main concept underlying `folk:Responsiveness` is the ability to promptly respond to a given input, which includes various forms of communication such as punchlines, puns, and witty answers. This example explicitly highlights the importance of the `folk:Responsiveness` value in everyday communication.

The second example, derived from Graph 2498, is extracted from the sentence: 'So Marine Le Pen says she wants to leave the EU to <u>protect national sovereignty</u> and break out of EU regulations [...].' This example demonstrates a more semantically intricate `mft:Care` situation. It is triggered by the word `wn:protect-verb-1`, with the Theme role consisting of the `fs:PoliticalLocales` frame (a trigger for `mft:Authority`), the term `wn:sovereignty-noun-1` (which triggers `folk:Independence` and `mft:Authority`), and the adjective `wn:national-adjective-1` (a trigger for `mft:Loyalty`). The value structure underlying this sentence involves the protection and support of a form of authority and independence, related to a specific in-group loyalty.

Betrayal Only a small number of matches between a `mft:Betrayal` situation and other values are found in MFRC, with a total of 11 occurrences across 4 graphs. None of these instances exhibit a particularly meaningful value structure. However, we present an example from Graph 2918, derived from the sentence: 'You don't know what happened because it never went to trial. The information you are interpreting could just as likely have had an agenda. This case was in the international limelight, so just because you live there doesn't give you any more authority than anyone else with what actually transpired. And Wikileaks didn't leak Macron's emails. That is fake news.' In this example, the occurrence of a `mft:Betrayal` situation is linked to the notion of *revealing a secret*, specifically private information. The Agent role in this situation is associated with another trigger for `mft:Betrayal`, namely `wn:bogus-adjectivesatellite-1`, and the `fs:Artificiality` frame. Although this example may not perfectly represent the epistemic stance or the underlying meaning that can be extracted from value semantic knowledge, it offers an opportunity to reflect on instances where diverse situations co-occur, evoking the same value frame. It becomes evident that further theoretical work is required to refine the granularity and semantics of `mft:Betrayal` situations, which can be more effectively operationalized during the detection phase.

Loyalty Moving on to the `mft:Loyalty` frame, it occurs in 104 cases across 46 graphs. An example of `mft:Loyalty` is illustrated in Graph 8393, which represents *working together for a common purpose*. The original text of this graph is: 'For the US, we need like, we're at the point where there needs to be WW2 level decimation of our population and infrastructure to get people to embrace universal healthcare.' In this example, the trigger for `mft:Loyalty` is `wn:embrace-verb-1`, where the Actor2 role pertains to an individual, resulting in the `wn:healthcare-noun-1` trigger for `mft:Care`.

Cheating The `mft:Cheating` frame is observed 185 times across 62 graphs. An interesting example is found in Graph 13,373, derived from the sentence: 'Your boss steals much more from you than taxes ever have.' In this case, the `mft:Cheating` situation is triggered by `vb:Steal_10050000`, which intriguingly assigns the role of Agent to the WordNet synset `wn:foreman-noun-1`, triggering the values `folk:Leadership`, `bhv:Power`, and `mft:Authority`. The value structure extracted from this sentence represents a source of power and authority engaging in an act of cheating. While we cannot directly derive an epistemic stance from this example, we can infer a latent conflict between the role of authority and its involvement as the Agent in a `mft:Cheating` situation.

Fairness Moving on to the mft:Fairness frame, it is encountered in 532 pattern matches across 185 MFRC graphs. Graph 11,177 provides an illustration of this frame in the sentence: 'What part of the catholic doctrine allows the cover up of widespread molestation of children?' In this example, the mft:Fairness trigger vb:Allow_65000000 designates the "catholic doctrine" as the Agent, which triggers folk:Belief and mft:Sanctity, while the fictive Location is represented by wn:harassment-noun-2. It may appear peculiar to have the harassment event as the Location rather than the Theme of the allowing event, but the actual Theme is the "cover up" situated metaphorically in the context of a harassment event.

Degradation Regarding the mft:Degratation frame, it appears in 199 matches across 66 graphs. The example provided here is from Graph 4,832, originating from the sentence: 'There is none. It's a non-paid job, and it's essentially what the US' First Lady does. There's no corruption, it's just Trump and Putin bots spewing salt that Le Pen didn't win.' The mft:Degratation situation is triggered by vb:Spew_43040000, where the Agent role is filled by entities that evoke the fs:Leadership frame. The value structure depicted in this sentence highlights an authority figure being the Agent of a mft:Degratation situation.

Sanctity The mft:Sanctity frame appears in 554 instances across 149 graphs. Graph 1,422 is generated from the sentence: 'Le Pen really needs a miracle, anyone comparing them really needs to look at the polls. She's praying for an extremely violent and bloody terrorist attack.' In this case, a mft:Sanctity situation is triggered by vb:Pray_32020100, with the Theme being a mft:Harm trigger disambiguated as wn:attack-noun-1 and wn:violent-adjective-1. This represents one of those ambiguous cases where a harmful event is perceived as positive from a certain perspective. While it is not possible to derive the epistemic stance solely from the evocation of the value, this method allows for the retrieval and investigation of such semantic dependency structures.

Oppression Moving on to the mft:Oppression value, it occurs in 397 situations across 149 graphs. Graph 11,656 provides a relevant example: 'Being a sex worker isn't demeaning in a world where capitalism forces people to make ends. Let's stop demeaning sex work Reddit by acting like it's slight above data entry bullshit.' In this instance, the concept of wn:capitalism-noun-1 acts as the Agent triggering an mft:Oppression situation through vb:Force_59000000. The significance lies in wn:capitalism-noun-1 being a trigger for the folk value folk:Structure, which embodies the idea of having a well-defined shape, solidity, and organization. This example highlights how a specific occurrence of a value can be perceived as unbearable from a particular perspective. One hypothesis regarding folk values is that, as primarily individual values shaping everyday behavior, they express their positive value charge most effectively by following an Aristotelian *ethic of the right middle*.

Authority Regarding the mft:Authority situation pattern, it is observed in 2,743 occurrences across 893 graphs. Graph 346 offers a good example: 'We won't elect Le Pen because she's a fascist lite, not because we're sexists.' In this example, the trigger for the mft:Authority situation is wn:elect-verb-1, while the DBpedia entity db:Fascism serves as the trigger for mft:Oppression. Furthermore, the graph also includes information about the negation of the mft:Authority situation, represented in FRED semantics as: mft:Authority :hasTruthValue :False.

Finally, concerning the MFT values, no meaningful matches are found for mft:Liberty and mft:Subversion. Thus, we proceed with the analysis using the BHV values.

Power Regarding the BHV values, bhv:Power situations occur in 1,454 instances across 435 graphs. Graph 3,744 is generated from the sentence: 'French politician who can lead a rebirth of the EU. I still don't know too much, but this sounds a bit like Macron.' This graph provides an example of the frequent co-occurrence of the bhv:Power frame and the mft:Authority frame. In this sentence, the trigger wn:politician-noun-2, which evokes mft:Authority, bhv:Power, and folk:Leadership, serves as the Agent of the bhv:Power situation triggered by wn:lead-verb-1.

Achievement Moving on to bhv:Achievement situations matching the desire pattern, there are 32 occurrences across 14 graphs. Graph 1,166 is an example generated from the sentence: 'I've always heard that regulatory protections for workers in Nordic countries are pretty 'weak' compared to the rest of Europe. This is what Macron is trying to achieve in France, and he's not perceived as a social democrat.' In this case, the Agent of the bhv:Achievement situation is the WordNet entity wn:Democrat-noun-1, which triggers the *power* cluster of values, including mft:Authority, bhv:Power, and folk:Leadership.

Benevolence For the bhv:Benevolence value, there are 157 matches for the desired pattern across 63 graphs. Graph 5,685 is generated from the sentence: 'They mean 30,000 pro Le Pen accounts to help the opposition win.' This example demonstrates how the bhv:Benevolence situation is triggered by the WordNet verb wn:help-verb-1, while its Beneficiary role, representing a "winning event," is filled by wn:win-noun-1, which triggers the folk:Victory value.

Security The SPARQL pattern for `bhv:Security` situations is matched in 516 occurrences across 154 graphs. Graph 11,179 provides an example with the sentence: 'You don't see how the desire to preserve 'normalcy' is exactly what protected the brother and prolonged the abuse?' In this sentence, the trigger for `bhv:Security` is `vb:Protect_85000000`, and its Theme role is filled by `wn:brother-noun-1`, which triggers the `folk:Family` value.

Regarding the remaining values, `bhv:Hedonism`, `bhv:Conformity`, `bhv:Tradition`, `bhv:SelfDirection`, `bhv:Stimulation`, and `bhv:Universalism`, no meaningful results were obtained. In the case of `bhv:Hedonism` and `bhv:Stimulation`, this could be attributed to the limited number of lexical and factual triggers retrieved by the QUOKKA workflow. For `bhv:Conformity`, `bhv:Tradition`, and `bhv:SelfDirection`, it is possible that these values are less easily detected from a single lexical trigger but could be retrieved through more specific patterns that involve multiple frames filling specific roles. The next section will focus on some experiments related to the compositional aspects of value frames.

5.2.6 Value Compositionality Patterns

In Chapter 3, we discussed how MFT universal values, in particular, are often defined in a circular manner, referring to the concept itself to define it. However, this is not necessarily true for all values. Schwartz recognizes the hierarchical organization of values as one of their main characteristics [297]. Based on this understanding, we propose graph patterns to detect specific combinations of values, building upon the specific retrievals made in the previous sections. These patterns can be utilized to identify values that have not yet been operationalized or to capture complex concepts by leveraging the semantic compositionality of image-schematic, emotion, and value frames.

Negation The first modifier we consider is the Truth value of a value. Negating a value does not automatically imply a commitment to its opposite pole (in MFT) or quadrant (in BHV), but it is still relevant for further inferences. In FRED, negations are treated by attaching the property `fred:hasTruthValue` followed by the value `fred:False`. As seen in the previous section, negating values like `bhv:Power` and `folk:Control` can result in the evocation of negative emotions.

Blockage Patterns The concept of BLOCKAGE, as observed in the previous section, carries a semantic notion of *proactive opposition*. Therefore, when encountering a BLOCKAGE situation, if the `is:BLOCKED` role is filled by a value trigger, the commitment or epistemic stance can vary significantly.

However, there are some caveats to consider. First, the `is:BLOCKER` entity should be either a Cognizer or a value trigger. The Blockage of `mft:Harm` can be seen as implying a (local) commitment to `mft:Care`, while the opposite is not necessarily true. Similarly, blocking `mft:Cheating` can qualify the `is:BLOCKER` as being committed to `mft:Fairness`, at least in that specific occurrence. This semantics aligns with the notion of MFT dyads, where values are considered in contrast to their violations. Therefore, a proactive opposition towards the negative pole of a dyad can be seen as an indirect commitment to the positive pole. However, the opposite does not hold true. Finally, when the `is:BLOCKER` entity is a value trigger, it can be said to have, if not a fully negative epistemic stance, at least a contrastive position against the `is:BLOCKED` entity.

Support Patterns In the case of a Support situation, which carries the semantics of proactive support, if the `is:SUPPORTER` role is filled by a Cognizer or a value trigger, it usually indicates a positive epistemic stance towards the `is:SUPPORTED` entity. These patterns enable the identification of complex relationships between values and facilitate the exploration of indirect commitments or epistemic stances within the value framework.

Harm Patterns Due to the inherent semantics of `mft:Harm`, where the Agent inflicts harm upon the Patient, it can be inferred that the Agent is considered dysphoric from the perspective of the Patient.

Care Patterns In contrast to the `mft:Harm` situation, any Agent in a `mft:Care` situation is presumed to have a positive epistemic stance towards the Patient/Theme of the `mft:Care`. From the Patient's perspective, the Agent is seen as someone who brings about a sense of euphoria or well-being.

Emotion patterns In the context of negative emotions (`be:Fear`, `be:Sadness`, `be:Disgust`, `be:Anger`), the Patient of such an emotion situation, with the Cause entity as the trigger, is likely to have a negative epistemic stance towards the Cause entity. Conversely, the Cause entity is regarded as dysphoric from the perspective of the Patient. These patterns provide insights into the nuanced relationships between values and emotions, shedding light on the epistemic stances and affective experiences associated with different situations.

5.2.7 Value Compositionality Testing

In this section, we present a series of knowledge extraction experiments based on the patterns proposed in the previous section. These experiments involve querying the MFRC

(Multilingual FrameNet Resource Construction) graphs using SPARQL. Our first example focuses on the BHV (Basic Human Values) model, specifically its 2019 version described by Giménez-García and colleagues [108]. The BHV model, illustrated in Figure 3.2 in Chapter 3, provides specifications for a more general version of the 10 BHV values, as operationalized in Section 3.2.3.

For instance, the value `bhv:Societal` is considered a subclass of `bhv:Security`, characterized by the attitudes `bhv:SocialFocus` and `bhv:SelfProtectionAndAnxietyAvoidance`. It specifically pertains to the security of a societal system or institution. Interestingly, the `bhv:Societal` value is not included in the BHV lexicon used for operationalization of the theory. Its lexical manifestation primarily relies on periphrasis and expressions of appraisal rather than specific lexical units. Consequently, to assess the validity of the compositionality expressed in the previous paragraph, we are conducting tests on those BHV values that are acknowledged to exist but have not yet been operationalized.

Considering other values and the availability of existing FrameNet frames, we have developed a graph pattern to extract the Societal value frame. This pattern is expressed as a SPARQL construct query, presented below:

```
CONSTRUCT { ?node1 vcvf:triggers bhv:Societal .  }
WHERE {
{ ?node1 rdf:type/owl:equivalentClass ?class1 .
?class1 vcvf:triggers bhv:Security .
?node1 role:Theme|role:Patient|
role:Undergoer|role:Experiencer ?node1 .
?node1 (<>|!<>)* fs:Organization .
}
UNION
{ ?node1 rdf:type/owl:equivalentClass ?class1 .
?class1 vcvf:triggers bhv:Security .
?node1 role:Theme|role:Patient|
role:Undergoer|role:Experiencer ?node1 .
?node1 (<>|!<>)* fs:Institutions . }
}
```

In the SPARQL construct query, we specify that nodes belonging to a certain class trigger the `bhv:Security` value, and we further require that these nodes have specific VerbNet roles (Theme, Undergoer, Patient, or Experiencer) towards entities of the class `fs:Institution` or `fs:Organization`. These triggering nodes, denoted as `?node1`, correspond to entities that activate the bhv:Societal value.

However, unfortunately, due to the limited size of the bhv:Security triggers graph, no instances of bhv:Societal were retrieved using this approach. This provides an opportunity for further testing. It can be argued that ?node1, instead of being exclusively of type bhv:Security, could also belong to other values such as mft:Care, bhv:Benevolence, folk:Helping, or even Support, as the meaning of the bhv:Societal value encompasses caring, protecting, and supporting institutions or organizations.

Moreover, we could broaden the scope by not specifically targeting the frames fs:Institutions and fs:Organization, but instead searching for generic triggers of mft:Authority. However, this would introduce some unnecessary noise into the results. To test our hypothesis, we repeated the query with mft:Care, bhv:Benevolence, and Support.

The experiment yielded positive results. Among several instances, Graph 3166 precisely matches the desired pattern for all the aforementioned variables. The sentence 'Good! Merkel/Macron drawing a strong example out of this will help with the stability of the EU as a legitimate supranational institution.' exemplifies this match. The VerbNet entity vb:Help_72000100 triggers the values mft:Care, Support, folk:Helping, and bhv:Benevolence in this graph. It assumes the Theme role with "European Union" as the evoked fs:Institutions frame.

Similarly, we can develop a pattern for the sibling class of bhv:Societal, which is bhv:Personal. In this case, the restriction should focus on entities that evoke the fs:People frame rather than the fs:Organization or fs:Institutions frames. Consequently, we modified the query accordingly and repeated the investigation. In this case, the investigation was successful by directly using bhv:Security as the main frame, resulting in Graph 2931. The graph is generated by the following sentence: 'Guess what? We don't vote by rule of law. Anybody can decide whatever the fuck they want based on the evidence they see. If you want to turn a blind eye to Putin's funding of Le Pen I guess that's your choice. It just makes you ignorant. And other people are gonna call you on it. And guess what, they don't have to form their opinion of you inside a courtroom either, since that seems to be the only place you care to defend people's actions.'

In this instance, the VerbNet entity vb:Defend_85000000 assumes the Theme role, with a node evoking the fs:People frame.

Considering another example, the 2019 version of BHV includes other more specific values, such as bhv:Action, which is a subclass of bhv:SelfDirection. The main idea behind this value is *the importance of freedom of action, expression, and movement*. We modified the previous pattern to search for situations related to bhv:SelfDirection, where some role is associated with several frames: fs:Opinion, fs:Telling, fs:Reasoning, and fs:ExpressingPublicly. However, no results were found using this modification. Therefore, we further modified the pattern to search for situations related to mft:Liberty, bhv:Benevolence, and Support. Finally, we achieved success with Graph 624, which contains the following sentence: 'Yes, so that is exactly why it is very hard to portray people as 'liberals always do that' and stuff because we are very different from the US with different values. We don't value your freedom of speech as much as you do. We don't like people throwing dangerous noise in public communication. We have a different approach to that. Also, if you can classify Macron as neoliberal, it will be different from neoliberals in the US. And I wouldn't say here that Macron would classify as an SJW when the majority of the country thinks the opposite.'

Finally, we tested one last hypothesis regarding compositionality: the ability to detect the transgression or violation of some non-formalized or lexicalized value.

Let's consider another 2019 BHV value: bhv:Rules, which is a subclass of bhv:Conformity. The main idea behind this value is to *follow the laws or rules of a certain social community*. We modified the query to retrieve mft:Harm situations, where an entity evokes the fs:Law frame.

Ultimately, we were able to retrieve several graphs, including Graph 3197, generated from the following sentence (shortened for brevity): 'That's where you're wrong, kiddo. Everyone voted for Trump thinking he wouldn't be that bad. He hasn't even elected any state officials yet. No wall. Self-proclaimed globalist. Hires an advisor who actively says he's trying to burn the government down.' The underlined part demonstrates how the wn:burn-verb-1, which triggers mft:Harm, takes as its Theme an entity that evokes the fs:Law frame, matching the desired pattern.

Therefore, social (and consequently legal) norms can be derived by substituting mft:Care with its violation, mft:Harm.

5.2.8 SPICE Data Enrichment

The value-emotion frame-based detection has been utilized within the context of the SPICE H2020 European Project to enhance cultural heritage data from various European cultural institutions [199]. Although the data itself is not open, as it is the property of individual cultural institutions, it is accessible through the SPICE Linked Data Hub (LDH).

The process of enriching this data serves two main purposes. Firstly, it enables the clustering of users based on the provided interpretations. This clustering can be utilized to group users who share similar values and emotions, or alternatively, to establish an interpretation-reflection loop [199]. This can be achieved by designing a recommender system that suggests artworks evoking emotions or values opposite to those that have been positively appraised by a particular user. The goal is to promote reflection and encourage the confrontation of diversities.

Secondly, from the perspective of cultural institutions, this annotation process allows for the meaningful organization of cultural items within available exhibition spaces. It facilitates the grouping of artworks that convey a similar emotional or value-based impact within the same room. Alternatively, the items can be organized to create a progression of emotional or value intensity. Additionally, artworks labeled with opposing or conflicting emotions or values can be categorized as controversial. Once this categorization is known, the cultural institution can handle such artworks according to its own preferences.

The Emotion-Value enrichment has been applied to various types of datasets obtained from different institutions, which will be described in the following paragraphs.

GAM - Turin Modern Art Gallery The value and emotion enrichment process was carried out on 56 paintings' descriptions obtained from the catalogue of the Modern Art Gallery in Turin. These artworks were also utilized in an annotation experiment involving the exhibition audience. Annotators were asked a series of questions for each painting, including: (i) 'how does it make you feel?', (ii) 'what does it remind you of?', (iii) 'what does it make you think about?', and (iv) 'if you had to give it a title, what would it be?'. The answers provided by the annotators, along with the descriptions of the items, were stored in a JSON file as key-value pairs.

Each of these answers, along with the corresponding item descriptions, were enriched with emotion and value knowledge. This resulted in a multi-perspective annotation, where the values and emotions evoked by the item's description are associated with the cultural artifact itself, while the values and emotions evoked by the annotator's answers are related to the individual perception and interpretation of the cultural artifact.

DMH - Design Museum of Helsinki The Design Museum of Helsinki conducted an experiment involving the collection of 406 user "stories", which are short texts inspired by 66 cultural artifacts from their collection. These stories were enriched with value and emotion content and then integrated into the SPICE Linked Data Hub, providing annotations about the user's interpretation of the artifacts.

While the detailed exploration of cultural data enrichment is beyond the scope of this work and the data is not openly available, it is a significant enrichment process and contributes substantially to the SPICE H2020 European project.

5.3 Chapter Conclusions

In this chapter, we conducted an experiment on 13,800 sentences from the Moral Foundations Reddit Corpus, utilizing the detectors described in the previous chapters. We enriched the resulting graphs, generated by the FRED tool, with image-schematic, value, and emotional knowledge. The purpose of this experiment was to explore the main relationships among image schemas, emotions, and values. We aimed to investigate their co-location and identify clusters organized based on the overlap in their semantic extension.

We demonstrated that these three levels of knowledge are interconnected by analyzing the semantic structure of individual frames, semantic roles, and the overlaps in their semantic trigger graphs. Additionally, we showcased how the resources developed in Chapters 2, 3, and 4 can be utilized to investigate the latent moral content (and not only moral) of natural language text. By leveraging the graph format and semantic relations, the detection process remains explainable and reproducible.

While our approach allows for a more profound exploration of semantic dependencies and the intertwined relationships of different levels of embodied, value, factual, and emotional knowledge, there is still room for considerable improvement to make this approach competitive with state-of-the-art classification systems.

All the inference patterns and queries presented in this chapter are available on the ValueNet GitHub[3] repository. They can be automatically executed to investigate any kind of text (transposed into a graph) and extract latent moral content. The automatic extraction of euphoric-dysphoric qualification and epistemic stance is performed using the patterns described in this chapter, while many others are still to be investigated. The code for the automatic detector of image schemas, values, and emotions is available and documented on the ValueNet GitHub.

[3]The documentation and graphs for the MFRC experiments are available here: https://github.com/StenDoipanni/ValueNet

Conclusions and Future Works

In this study, we embarked on an exploration of moral, social, and individual values using a grounded approach rooted in frame semantics and embodied cognition. To formalize the elusive concept of "value," we initially considered it as a semantic frame. Our approach involved adopting an ontological framework to formalize domain knowledge and enable automated inferences. Drawing upon prominent psychological and social theories that offer taxonomic organization systems for values, we translated them into ontological modules using the XD methodology and Ontology Design Pattern reuse. Additionally, we ensured alignment with the DOLCE ontology foundation.

Treating the notion of value as a frame allowed us to model multiple concurrent theories, considering each of them as an ontological module in its own right. By adopting this approach, various understandings of "value" from different theories were treated as subframes of a general notion of value, serving as a superclass encompassing all the previous ones. Consequently, different aspects of value semantics were captured in distinct ontological modules, facilitating the application of multiple theories for automated inferences.

Viewing a value as a semantic frame also prompted us to question the frame's structure and lexicalization. Leveraging the QUOKKA (Querying Ontological Knowledge for Knowledge Augmentation) workflow, we populated the graph of each value within every considered theory with semantic triggers. Unlike the lexical triggers used in FrameNet frames, our triggers were not limited to lexicographic data. Instead, we extracted them from the Framester ontology hub, which, through a network of alignments, facilitated the creation of a knowledge base comprising value triggers sourced entirely from Linked Open Data entities found in reputable semantic web resources such as WordNet, VerbNet, FrameNet, DBpedia, BabelNet, ConceptNet, and others.

Having populated the value graphs, we constructed a comprehensive frame-based value detector to evaluate the coverage and efficacy of our method. The developed detector leveraged the existing FRED tool, capable of generating graphs from natural language sentences, detecting frame evocations, performing lexical disambiguation using WordNet and VerbNet resources, and identifying factual knowledge from DBpedia.

The detector, utilizing the FRED-generated graphs, executes SPARQL queries against the Framester resource to search for a value activator among the graph nodes. Upon locating it, the detector attaches the corresponding triple, declaring the triggering value entity. However, due to the complex and socially negotiated nature of the value concept, which intertwines common-sense knowledge, individual perceptions, social constructs, moral ethics, and emotional appraisals, the ValueNet ontology alone is insufficient. It requires the incorporation of two additional levels of knowledge: (i) embodied cognition, which encompasses the conceptualization of values based on sensorimotor perception of the world, and (ii) the emotional dimension frequently associated with value situations.

To address these aspects, the ValueNet ontology is complemented by the ISAAC and EmoNet ontologies. ISAAC, the Image Schema Abstraction And Cognition ontology, serves as an encyclopedic resource on image schemas, which are sense-motor perception schemes that shape our conceptualization of the world. Developed following sound modeling practices, ISAAC treats each image schema as a frame and populates the graphs with semantic triggers sourced from a controlled vocabulary and semantic web resources.

On the other hand, the EmoNet ontology aims to model the domain of emotions, incorporating major state-of-the-art theories. Currently, EmoNet consists of a single theoretical module that formalizes Ekman's Basic Emotions theory into ontological form.

These three levels of knowledge, represented by the ValueNet, ISAAC, and EmoNet ontologies, are integrated into a single detector capable of identifying sensorimotor cognitive patterns, values, and emotions. To test potential inferences, experiments were conducted individually for each ontology network and overall to detect relevant patterns.

The subsequent chapters of this work provide a description of the content, organized as follows.

Chapter 1 provided the necessary theoretical and technical context for the subsequent chapters, outlining the main objective of this work: the formalization of moral, social, and individual values using a frame semantics approach grounded in embodied cognition. This chapter introduced the key semantic web resources utilized in the study, including the Framester ontology hub, the QUOKKA workflow for knowledge graph population, and the FRED tool for generating knowledge graphs from natural language.

In Chapter 2 focused on the domain of image schemas. It detailed the formalization of prominent image schema theories into the Image Schema Abstraction And Cognition (ISAAC) ontology. ISAAC consists of various modules, including J87, MPC, and HED, which are modeled based on literature, as well as ISCAT and ISFRAME, which capture image schematic cognitive metaphors. The ImageSchemaNet module operationalizes the ontology by declaring semantic triggers sourced from existing semantic web resources aligned within the Framester ontology. This module serves as a knowledge base of semantic triggers, enabling the development of an automatic tool for detecting image schemas from natural language using an explainable graph-based method. The chapter concludes with an evaluation of the ontologies' validity and the performance of the image schema detector.

Chapter 3 delved into the core modules of the study, namely universal moral values based on Moral Foundations Theory, socio-cultural values rooted in Basic Human Values theory, value situations represented as frames composed of "moral atoms" from Moral Molecules theory, and individual values extracted from the web using a bottom-up "folksonomy" approach. Each theory was formalized as an ontological module and imported into the ValueNet ontology network. Semantic triggers were then applied to operationalize each module, obtained from a controlled vocabulary derived from relevant literature. The chapter concluded with inference testing and experiments on value detection, showcasing the extraction of latent moral content from natural language using an explainable graph-based detector.

Chapter 4 focused on the emotional dimension of values, formalizing the Basic Emotion's theory and operationalizing it through a knowledge base of semantic triggers. The chapter concluded with experiments on emotion extraction from natural language using an automatic graph-based emotion detector.

the concluding chapter, consolidates the knowledge layers established in the preceding chapters by conducting an experiment on 13,800 sentences from the Moral Foundations Reddit Corpus (MFRC). It highlights the limitations of mere detection with single label annotations in explicitly capturing the moral content of certain sentences. In contrast, the graph-based tool developed in this work can detect image schematic knowledge, moral values, and emotions while preserving the semantic dependencies and role structure of the sentence. This graph-based structure enables the detection of combinations of embodiment, moral knowledge, and emotional appraisal, facilitating inferences regarding polarity attribution to a specific entity from the perspective of a cognizer and/or its epistemic stance about that entity.

The detection of these three layers is performed for each of the 13,800 sentences, enriching the graph with the corresponding levels of knowledge. Once the enriched graph is generated, the knowledge contained within it is explored using SPARQL queries, and several search patterns are illustrated.

The subsequent section of the chapter provides a summary of the modules presented throughout the thesis, along with any new material introduced. Additionally, various types of results are presented, which, although not the main focus of this work, are significant in their own right, although they are not described as supporting ontological tools or modules.

Summary of the Main Contributions

Throughout this work, we have presented a comprehensive set of ontology modules with diverse purposes. The following list provides an overview of these modules:

- A total of 20 ontological modules were developed, encompassing:
  - 7 modules focused on formalizing Image Schemas, including 4 modules dedicated to image schema theories and semantic triggers, 2 modules addressing cognitive metaphors, and 1 module transposing the Image Schema Language into OWL 2 syntax. All of these modules are imported into the comprehensive ISAAC (Image Schema Abstraction And Cognition) module, which aims to serve as an encyclopedic repository of multimodal knowledge related to image schemas.
  - 4 modules dedicated to emotions and appraisal theories, which are integrated into the EmoNet ontology to create an Atlas of Emotion theories.
  - 8 modules specifically devoted to values, representing different theories and their operationalization. These modules are imported into the ValueNet ontology network, which serves as a framework for formally representing major value theories and their lexical coverage. This allows for the detection of values from natural language using multiple theories simultaneously.
  - 1 module developed to facilitate the theoretical comparison of competing theories.
- Over 125,000 triples have been declared in these ontology modules.
- More than 20,000 graphs have been automatically generated from natural language sentences across six datasets. These graphs have been enriched with knowledge related to image schemas, emotions, and values.
- Three automatic detectors have been implemented, which can function individually or as three levels superimposed on each other. These detectors allow for the querying of knowledge through the SPARQL language, enabling the identification of patterns within the graph structure.

The resources developed throughout this work are still undergoing refinement and enrichment. As indicated in each section dedicated to the ontological modules, they are considered a work in progress with the aim of integrating numerous theoretical frameworks and multimodal data. This integration enables not only textual analysis but also the extraction of latent moral content from visual, sensorimotor, and behavioral perspectives.

5.3.1 By-Products

This section focuses on specific modules, tools, and resources that were developed for a particular purpose within this work but have potential applications beyond its scope.

- **QUOKKA:** The Querying Ontological Knowledge for Knowledge Augmentation workflow is a reusable tool designed to allow lexicographers, ontology engineers, and enthusiasts of web semantics to retrieve entities from well-known semantic web resources and reuse them in their projects. Originally developed as a "frame building workflow," it facilitates the retrieval of entities aligned in Framester while employing a frame semantics approach to model a specific domain. However, it can be utilized to gather knowledge for any purpose. The online user interface of QUOKKA enables individuals without SPARQL knowledge to retrieve data from Framester in a simple and user-friendly manner. The complete SPARQL queries are also available in the GitHub repository, allowing experts to fine-tune the search according to their specific needs or replicate the queries on the Framester endpoint. The QUOKKA workflow user interface can be accessed online[4].

- The `mft:Harm` semantic triggers graph, which is essentially an ontology in its own right, enables the identification of instances of physical or verbal violence. With appropriate heuristics and refinements, it could be utilized for tasks such as hate speech recognition or the identification of various forms of violence.

- The **MFRC graph**, which involves using FRED to generate a graph for each sentence in the Moral Foundations Reddit Corpus (MFRC), offers a comprehensive resource transposed into graph format. It is available online and enriched with values, emotions, and image schemas. Future research could utilize this knowledge base as a starting point to compare different approaches, develop hybrid methods, or improve upon the initial baseline.

- **ISL2OWL**, the ontological transposition of the Image Schema Language into OWL 2 syntax, was not extensively described in this work since it is not directly employed. However, it is currently being used and further developed in cognitive robotics experiments to explore image-schematic reasoning approaches for solving non-trivial Sokoban-like problems.

[4]The QUOKKA user interface is available here: `http://etna.istc.cnr.it/quokka/`

5.3.2 Future Works

The main ontology networks introduced in this work, namely ISAAC, ValueNet, and EmoNet, strive to be comprehensive knowledge repositories for the domains they model. As modular frameworks, they are open to enrichment and the addition of new theoretical modules and semantic triggers.

The proposed ontological resources are an ongoing process and can be improved in several aspects. Firstly (i), the coverage of theories within ISAAC and ValueNet currently includes the major existing theories on image schemas and values. However, there are still theories that have not been included and need to be supplemented to enhance the theoretical and conceptual coverage of the resources. Similarly, the EmoNet ontology currently lacks many of the most recent emotional models and theories, which are expected to be integrated in the near future.

The second aspect (ii) pertains to the cleaning of semantic trigger graphs. While the QUOKKA workflow facilitates knowledge extraction from the Framester resource, the subtle conceptual nuances of a given domain can often be best captured through careful manual selection and data cleaning.

The third aspect (iii) involves the introduction of multi-modal data to establish a robust knowledge base for the ontologies. Currently, linguistic resources have been utilized, where a graph is created for each text string and subsequently enriched with image schemas, values, and emotions. The objective is to also integrate other types of data, such as visual and numerical data resulting from experiments on sensorimotor perception, spatial coordinates of neural activation following specific inputs, and more.

By addressing these aspects, the ontological resources can continue to evolve and provide a more comprehensive and accurate representation of the domains they model.

The detection process of the current detector can be enhanced in several ways. Firstly (i), the transposition from natural language to a graph, which is the primary bottleneck, can benefit from improvements. Accurate detection of frames, precise disambiguation of verbs, nouns, and adjectives using resources like WordNet, VerbNet, FrameNet, etc., and proper identification of semantic dependencies by associating VerbNet roles correctly are crucial for improving the quality of knowledge extraction.

Secondly, hybridization of graph-based techniques with deep learning techniques and BERT-based models could be explored. This integration can introduce a "quantitative" measure to the detection process, such as a score indicating the activation intensity of a particular frame. This quantitative measure can provide additional information for more accurate detection on the graph.

Furthermore, considering the salience of the detected segment within a sentence can be an important parameter for improvement. In the case of longer textual chunks, the syntactic position or relevance of a specific element within the sentence may influence its salience score. This salience score can reflect the greater or lesser value/emotional load associated with that segment of the sentence.

By addressing these areas of improvement, the detector can enhance its performance in knowledge extraction, providing more accurate and nuanced results.

There are two operational improvements that are currently being tested.

The first improvement involves the use of a tool called Sentilo, which is derived from FRED. Sentilo assigns a polarity score to terms in a sentence by utilizing resources such as SentiWordNet and WordNet Affect. It incorporates verb semantics modeled after Levin's classification. For instance, the verb "to kill" is represented with two main parameters: the Agent of kill is associated with a negative epistemic stance from a socio-cultural perspective, while the Patient of kill is identified as the object of a negative epistemic stance. These semantics are then transposed into SPARQL queries that consider the epistemic stance of specific verbs, and they are integrated into the activation patterns discussed earlier.

The second improvement being tested involves an enhanced version of FRED based on the Abstract Meaning Representation (AMR) architecture, which represents a state-of-the-art approach. Compared to the current version of FRED, the AMR architecture not only improves frame detection but also enhances the identification of semantic dependencies, resulting in a more accurate and "clean" graph representation. An essential feature of the AMR architecture is its ability to process input text strings in different languages. Unlike the current FRED-based detector, which only works for English and requires translation for non-English inputs, the AMR architecture enables direct input of sentences in their original language.

These improvements aim to enhance the performance and functionality of the detection process, providing more accurate and efficient knowledge extraction capabilities.

Bibliography

[1] Jennifer L. Aaker and Angela Y. Lee. “I” seek pleasures and “we” avoid pains: The role of self-regulatory goals in information processing and persuasion. *Journal of Consumer Research*, 28(1):33–49, 2001.

[2] Pierre Ageron. Le partage des dix-sept chameaux et autres exploits arithmétiques attribués à l’imâm Alî. mouvance et circulation de récits de la tradition musulmane chiite. *Revue d’Histoire des Mathématiques*, pages 1–41, 2013.

[3] F. Alemanno, E. Houdayer, M. Cursi, S. Velikova, M. Tettamanti, G. Comi, S.F. Cappa, and L. Leocani. Action-related semantic content and negation polarity modulate motor areas during sentence reading: An event-related desynchronization study. *Brain research*, 1484:39–49, 2012.

[4] Thomas Aquinas. *Summa Theologiae (1265-1273)*. John Mortensen and Enrique Alarcón, 1911.

[5] Oscar Araque, Lorenzo Gatti, and Kyriaki Kalimeri. Moral strength: Exploiting a moral lexicon and embedding similarity for moral foundations prediction. *Knowledge-based systems*, 191:105184, 2020.

[6] Luigi Asprino, Luana Bulla, Stefano De Giorgis, Aldo Gangemi, Ludovica Marinucci, and Misael Mongiovì. Uncovering values: Detecting latent moral content from natural language with explainable and non-trained methods. In *Proceedings of Deep Learning Inside Out (DeeLIO 2022): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*, pages 33–41, 2022.

[7] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. Dbpedia: A nucleus for a web of open data. In *The semantic web*, pages 722–735. Springer, 2007.

[8] Collin Baker. Framenet, present and future. In *The First International Conference on Global Interoperability for Language Resources*, pages 12–17, 2008.

[9] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics, 1998.

[10] Alexandra Balahur, Jesus M. Hermida, and Andres Montoyo. Building and exploiting emotinet, a knowledge base for emotion detection based on the appraisal theory model. *IEEE transactions on affective computing*, 3(1):88–101, 2011.

[11] Christa Baldauf. *Metapher und Kognition: Grundlagen einer neuen Theorie der Alltagsmetapher*. Lang, 1997.

[12] Anat Bardi, Rachel M. Calogero, and Brian Mullen. A new archival approach to the study of values and value–behavior relations: validation of the value lexicon. *Journal of Applied Psychology*, 93(3):483, 2008.

[13] Anat Bardi and Shalom H. Schwartz. Values and behavior: Strength and structure of relations. *Personality and social psychology bulletin*, 29(10):1207–1220, 2003.

[14] Lawrence W. Barsalou. Perceptual symbol systems. *Behavioral and brain sciences*, 22(4):577–660, 1999.

[15] Lawrence W. Barsalou and Christopher R. Hale. Components of conceptual representation: From feature lists to recursive frames. 1993.

[16] Frederic Charles Bartlett. A theory of remembering. *Remembering: A study in experimental and social psychology*, pages 197–214, 1932.

[17] Brandon Bennett and Claudia Cialone. Corpus guided sense cluster analysis: a methodology for ontology development (with examples from the spatial domain). In *FOIS*, pages 213–226, 2014.

[18] Jeremy Bentham and John Stuart Mill. *Utilitarianism and other essays*. Penguin UK, 2004.

[19] Tim Berners-Lee, James Hendler, Ora Lassila, Brian C. Chaboyer, K.C. Nicolaou, Christopher N.C. Boddy, W. Wayt Gibbs, Steven Ashley, Scott O. Lilienfeld, and James M. Wood. 34 the semantic web. 2001.

[20] John W. Berry, Ype H. Poortinga, Janek Pandey, Marshall H. Segall, and Çiğdem Kâğıtçıbaşı. *Handbook of cross-cultural psychology: Theory and method*, volume 1. John Berry, 1997.

[21] Franck Berthelon and Peter Sander. Emotion ontology for context awareness. In *2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 59–64. IEEE, 2013.

[22] Tarek R. Besold, Maria M. Hedblom, and Oliver Kutz. A narrative in three acts: Using combinations of image schemas to model events. *Biologically inspired cognitive architectures*, 19:10–20, 2017.

[23] Arianna Betti and Hein Van Den Berg. Modelling the history of ideas. *British Journal for the History of Philosophy*, 22(4):812–835, 2014.

[24] Douglas Biber and Edward Finegan. Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text-interdisciplinary journal for the study of discourse*, 9(1):93–124, 1989.

[25] Wolfgang Bilsky and Shalom H. Schwartz. Values and personality. *European journal of personality*, 8(3):163–181, 1994.

[26] Eva Blomqvist, Valentina Presutti, Enrico Daga, and Aldo Gangemi. Experimenting with extreme design. In *International Conference on Knowledge Engineering and Knowledge Management*, pages 120–134. Springer, 2010.

[27] Stefano Borgo, Roberta Ferrario, Aldo Gangemi, Nicola Guarino, Claudio Masolo, Daniele Porello, Emilio M. Sanfilippo, and Laure Vieu. Dolce: A descriptive ontology for linguistic and cognitive engineering. *Applied Ontology*, (Preprint):1–25, 2022.

[28] Pierre Bourdieu. *A social critique of the judgement of taste*. Traducido del francés por R. Nice. Londres, Routledge, 1984.

[29] Robert Boyd and Peter J. Richerson. *The origin and evolution of cultures*. Oxford University Press, 2005.

[30] Margaret M. Bradley, Mark K. Greenwald, Margaret C. Petry, and Peter J. Lang. Remembering pictures: pleasure and arousal in memory. *Journal of experimental psychology: Learning, Memory, and Cognition*, 18(2):379, 1992.

[31] Marilynn B. Brewer and Ya-Ru Chen. Where (who) are collectives in collectivism? toward conceptual clarification of individualism and collectivism. *Psychological review*, 114(1):133, 2007.

[32] George Bruseker, Nicola Carboni, and Anaïs Guillem. Cultural heritage data management: the role of formal ontology and cidoc crm. *Heritage and archaeology in*

the digital age: acquisition, curation, and dissemination of spatial cultural heritage data, pages 93–131, 2017.

[33] Sven Buechel and Udo Hahn. Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. *arXiv:2205.01996*, 2022.

[34] Robert L. Campbell, John Chambers Christopher, and Mark H. Bickhard. Self and values: An interactivist foundation for moral development. *Theory & Psychology*, 12(6):795–823, 2002.

[35] Lea Canales and Patricio Martínez-Barco. Emotion detection from text: A survey. In *Proceedings of the workshop on natural language processing in the 5th information systems research working days (JISIC)*, pages 37–43, 2014.

[36] Walter B. Cannon. The james-lange theory of emotions: A critical examination and an alternative theory. *The American journal of psychology*, 39(1/4):106–124, 1927.

[37] Cristiano Castelfranchi. Il valore degli scopi: non solo duplice ma ibrido. *Sistemi intelligenti*, 34(3):471–488, 2022.

[38] Robert B. Cialdini, Raymond R. Reno, and Carl A. Kallgren. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*, 58(6):1015, 1990.

[39] Alan Cienki. Some properties and groupings of image schemas. *Lexical and syntactical constructions and the construction of meaning*, pages 3–15, 1997.

[40] Scott Clifford and Jennifer Jerit. How words do the work of politics: Moral foundations theory and the debate over stem cell research. *The Journal of Politics*, 75(3):659–671, 2013.

[41] Sara Coppini, Stefano De Giorgis, and Aldo Gangemi. framECO: an ontology to model the emotional experience in literary texts. In *Master Thesis in Digital Humanities and Digital Knowledge, University of Bologna*, 2022.

[42] Oliver Scott Curry, Mark Alfano, Mark J. Brandt, and Christine Pelican. Moral molecules: Morality as a combinatorial system. *Review of Philosophy and Psychology*, pages 1–20, 2021.

[43] Oliver Scott Curry, Matthew Jones Chesters, and Caspar J. Van Lissa. Mapping morality with a compass: Testing the theory of 'morality-as-cooperation'with a new questionnaire. *Journal of Research in Personality*, 78:106–124, 2019.

[44] Tim Dalgleish. The emotional brain. *Nature Reviews Neuroscience*, 5(7):583–589, 2004.

[45] Antonio Damasio. The neurobiological grounding of human values. In *Neurobiology of human values*, pages 47–56. Springer, 2005.

[46] Antonio R. Damasio. Descartes error revisited. *Journal of the History of the Neurosciences*, 10(2):192–194, 2001.

[47] Antonio R. Damasio. *Looking for Spinoza: Joy, sorrow, and the feeling brain*. Houghton Mifflin Harcourt, 2003.

[48] Oana David, Ellen Dodge, J. Hong, Elise Stickles, and E. Sweetser. Building the metanet metaphor repository: The natural symbiosis of metaphor analysis and construction grammar. In *The 8th International Construction Grammar Conference (ICCG 8)*, pages 3–6, 2014.

[49] Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. Automated hate speech detection and the problem of offensive language. In *Proceedings of the international AAAI conference on Web and Social Media*, volume 11, pages 512–515, 2017.

[50] Stefano De Giorgis and Aldo Gangemi. Exuviae: an ontology for conceptual epistemic comparison. In *Graphs and Networks in the Humanities 2022 - Proceedings of the 6th International Conference*, Amsterdam, Netherlands, 2022. Accepted for publication, forthcoming.

[51] Stefano De Giorgis, Aldo Gangemi, and Rossana Damiano. Basic human values and moral foundations theory in valuenet ontology. In *International Conference on Knowledge Engineering and Knowledge Management*, pages 3–18. Springer, 2022.

[52] Stefano De Giorgis, Aldo Gangemi, and Dagmar Gromann. ImageSchemaNet: Formalizing embodied commonsense knowledge providing an image-schematic layer to framester. *Semantic Web Journal*, 2022.

[53] Stefano De Giorgis, Aldo Gangemi, and Dagmar Gromann. Introducing ISAAC: the Image Schema Abstraction And Cognition Modular Ontology. In *Proceedings of the Joint Ontology Workshops 2022*, CEUR Workshop Proceedings. CEUR-WS.org, 2022.

[54] Jan De Houwer and Dirk Hermans. Cognition and emotion: Reviews of current research and theories. 2010.

[55] Frans B.M. De Waal and Frans de Waal. *Good natured: The origins of right and wrong in humans and other animals*. Number 87. Harvard University Press, 1996.

[56] Michael DeBellis. The Ethics Ontology: A Universal Moral Grammar. Published on Academia: https://www.academia.edu/15327812.

[57] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407, 1990.

[58] Morteza Dehghani, Kate Johnson, Joe Hoover, Eyal Sagi, Justin Garten, Niki Jitendra Parmar, Stephen Vaisey, Rumen Iliev, and Jesse Graham. Purity homophily in social networks. *Journal of Experimental Psychology: General*, 145(3):366, 2016.

[59] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. Goemotions: A dataset of fine-grained emotions. *arXiv:2005.00547*, 2020.

[60] Julien Deonna and Fabrice Teroni. *The emotions: A philosophical introduction*. Routledge, 2012.

[61] Oxford English Dictionary. Oxford english dictionary. *Simpson, Ja & Weiner, Esc*, 3, 1989.

[62] Rim Djedidi and Marie-Aude Aufaure. Onto-evo a l an ontology evolution approach guided by pattern modeling and quality evaluation. In *International Symposium on Foundations of Information and Knowledge Systems*, pages 286–305. Springer, 2010.

[63] Ellen K. Dodge, Jisup Hong, and Elise Stickles. Metanet: Deep semantic automatic metaphor analysis. In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 40–49, 2015.

[64] Otniel E. Dror. The cannon–bard thalamic theory of emotions: A brief genealogy and reappraisal. *Emotion Review*, 6(1):13–20, 2014.

[65] Emmanuel Dupoux and Pierre Jacob. Universal moral grammar: A critical appraisal. *Trends in cognitive sciences*, 11(9):373–378, 2007.

[66] Astrid Duque-Ramos, Jesualdo Tomás Fernández-Breis, Robert Stevens, and Nathalie Aussenac-Gilles. Oquare: A square-based approach for evaluating the quality of ontologies. *Journal of research and practice in information technology*, 43(2):159–176, 2011.

[67] Abraham Edel. The concept of value and its travels in twentieth-century america. *Values and Value Theory in Twentieth-Century America: Essays in Honor of Elizabeth Flower*, pages 12–36, 1988.

[68] Paul Ekman. Basic emotions. *Handbook of cognition and emotion*, 98(45-60):16, 1999.

[69] Paul Ekman. Atlas of emotion, 2018. Online: https://atlasofemotions.org/.

[70] Dylan Evans. *Emotion: A very short introduction*. OUP Oxford, 2003.

[71] Gilles Fauconnier. *Mental spaces: Aspects of meaning construction in natural language*. Cambridge University Press, 1994.

[72] Gilles Fauconnier and Mark B. Turner. Blending as a central process of grammar: Expanded version. *Conceptual Structure, Discourse, and Language, Adele Goldberg, ed*, pages 113–130, 1998.

[73] Paul Feyerabend. *Science in a Free Society*. Verso Books, 2018.

[74] Charles J. Fillmore. Frame semantics. In *Linguistics in the Morning Calm*, pages 111–138. Seoul: Hanshin, 1982.

[75] Charles J. Fillmore. Frames and the semantics of understanding. *Quaderni di semantica*, 6(2):222–254, 1985.

[76] John Finnis. *Natural law and natural rights*. Oxford University Press, 2011.

[77] Ludwik Fleck. *Entstehung und Entwicklung einer wissenschaftlichen Tatsache Einführung in die Lehre vom Denkstil und Denkkollektiv*. B. Schwabe, 1935.

[78] Nico H. Frijda et al. *The emotions*. Cambridge University Press, 1986.

[79] Jeremy A. Frimer, Reihane Boghrati, Jonathan Haidt, Jesse Graham, and Morteza Dehgani. Moral Foundations Dictionary for Linguistic Analyses 2.0. *Unpublished manuscript*, 2019. Online: https://osf.io/ezn37/.

[80] Dean Fulgoni, Jordan Carpenter, Lyle Ungar, and Daniel Preoţiuc-Pietro. An empirical exploration of moral foundations theory in partisan news sources. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3730–3736, 2016.

[81] Hans-Georg Gadamer. *Truth and Method*. 1st German ed., 1960, 1st English ed. 1975, Continuum / Seabury Press, 2004. trans. by Weinsheimer, J., Marshall D.G.

[82] Vittorio Gallese. The 'shared manifold' hypothesis. from mirror neurons to empathy. *Journal of consciousness studies*, 8(5-6):33–50, 2001.

[83] Vittorio Gallese. Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the cognitive sciences*, 4(1):23–48, 2005.

[84] Vittorio Gallese. Mirror neurons and the social nature of language: The neural exploitation hypothesis. *Social neuroscience*, 3(3-4):317–333, 2008.

[85] Vittorio Gallese, Christian Keysers, and Giacomo Rizzolatti. A unifying view of the basis of social cognition. *Trends in cognitive sciences*, 8(9):396–403, 2004.

[86] Vittorio Gallese and George Lakoff. The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive neuropsychology*, 22(3-4):455–479, 2005.

[87] Antony Galton. The Formalities of Affordance. In Mehul Bhatt, Hans Guesgen, and Shyamanta Hazarika, editors, *Proceedings of ECAI2010 Workshop on Spatio-Temporal Dynamics*, pages 1–6, 2010.

[88] Aldo Gangemi. Norms and plans as unification criteria for social collectives. *Autonomous Agents and Multi-Agent Systems*, 17(1):70–112, 2008.

[89] Aldo Gangemi. Closing the loop between knowledge patterns in cognition and the semantic web. *Semantic Web*, 11(1):139–151, 2020.

[90] Aldo Gangemi, Mehwish Alam, Luigi Asprino, Valentina Presutti, and Diego Reforgiato Recupero. Framester: a wide coverage linguistic linked data hub. In *European Knowledge Acquisition Workshop*, pages 239–254. Springer, 2016.

[91] Aldo Gangemi, Mehwish Alam, and Valentina Presutti. Amnestic forgery: An ontology of conceptual metaphors. In Stefano Borgo, Pascal Hitzler, and Oliver Kutz, editors, *Proceedings of the 10th International Conference, FOIS 2018*, volume 306 of *Frontiers in Artificial Intelligence and Applications*, pages 159–172. IOS Press, 2018.

[92] Aldo Gangemi, Carola Catenacci, Massimiliano Ciaramita, Jos Lehmann, Rosa Gil, Francesco Bolici, and Onofrio Strignano. Ontology evaluation and validation. *An integrated formal model for the quality diagnostic task*, 30:36, 2005.

[93] Aldo Gangemi and Dagmar Gromann. Analyzing the imagistic foundation of framality via prepositions. In Adrien Barton, Selja Seppälä, and Daniele Porello, ed-

itors, *Proceedings of the Joint Ontology Workshops 2019*, volume 2518 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2019.

[94] Aldo Gangemi, Nicola Guarino, Claudio Masolo, and Alessandro Oltramari. Sweetening WordNet with DOLCE. *AI magazine*, 24(3):13–13, 2003.

[95] Aldo Gangemi and Peter Mika. Understanding the semantic web through descriptions and situations. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 689–706. Springer, 2003.

[96] Aldo Gangemi and Valentina Presutti. Ontology design patterns. In *Handbook on ontologies*, pages 221–243. Springer, 2009.

[97] Aldo Gangemi, Valentina Presutti, and Mehwish Alam. Amnestic forgery: An ontology of conceptual metaphors. In *Formal Ontology in Information Systems - Proceedings of the 10th International Conference, FOIS 2018, Cape Town, South Africa, 19-21 September 2018*. IOS Press, September 2018.

[98] Aldo Gangemi, Valentina Presutti, Diego Reforgiato Recupero, Andrea Giovanni Nuzzolese, Francesco Draicchio, and Misael Mongiovì. Semantic web machine reading with FRED. *Semantic Web*, 8(6):873–893, 2017. 10.3233/SW-160240.

[99] Aldo Gangemi, Valentina Presutti, Diego Reforgiato Recupero, Andrea Giovanni Nuzzolese, Francesco Draicchio, and Misael Mongiovì. Semantic web machine reading with FRED. *Semantic Web*, 8(6):873–893, 2017.

[100] Justin Garten, Reihane Boghrati, Joe Hoover, Kate M. Johnson, and Morteza Dehghani. Morality between the lines: Detecting moral sentiment in text. In *Proceedings of IJCAI 2016 workshop on Computational Modeling of Attitudes*, 2016.

[101] R. Gatti, A. Tettamanti, P.M. Gough, E. Riboldi, L. Marinoni, and G. Buccino. Action observation versus motor imagery in learning a complex motor task: a short review of literature and a kinematics study. *Neuroscience letters*, 540:37–42, 2013.

[102] Diman Ghazi, Diana Inkpen, and Stan Szpakowicz. Detecting emotion stimuli in emotion-bearing sentences. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 152–165. Springer, 2015.

[103] Raymond W. Gibbs and Raymond W. Gibbs. *Intentions in the Experience of Meaning*. Cambridge University Press, 1999.

[104] Raymond W. Gibbs Jr., Dinara A. Beitel, Michael Harrington, and Paul E. Sanders. Taking a stand on the meanings of stand: Bodily experience as motivation for polysemy. *Journal of Semantics*, 11(4):231–251, 1994.

[105] James J. Gibson. The theory of affordances, in perceiving, acting, and knowing. towards an ecological psychology. In Robert Shaw and John Bransford, editors, *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, pages 67–82. NJ: Lawrence Erlbaum, Hillsdale, 1977.

[106] Gerd Gigerenzer and Peter M. Todd. Fast and frugal heuristics: The adaptive toolbox. In *Simple heuristics that make us smart*, pages 3–34. Oxford University Press, 1999.

[107] Carol Gilligan. New maps of development: new visions of maturity. *American Journal of Orthopsychiatry*, 52(2):199, 1982.

[108] August Corrons Giménez and Lluís Garay Tamajón. Analysis of the third-order structuring of shalom schwartz's theory of basic human values. *Heliyon*, 5(6), 2019.

[109] Joseph Grady. The "conduit metaphor" revisited: A reassessment of metaphors for communication. *Discourse and cognition: Bridging the gap*, pages 205–218, 1998.

[110] Jesse Graham, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P. Wojcik, and Peter H. Ditto. Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology*, volume 47, pages 55–130. Elsevier, 2013.

[111] Jesse Graham, Jonathan Haidt, and Brian A. Nosek. Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, 96(5):1029, 2009.

[112] Jesse Graham, Brian A. Nosek, and Jonathan Haidt. The moral stereotypes of liberals and conservatives: Exaggeration of differences across the political spectrum. *PloS one*, 7(12):e50092, 2012.

[113] Jesse Graham, Brian A. Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva, and Peter H. Ditto. Mapping the moral domain. *Journal of personality and social psychology*, 101(2):366, 2011.

[114] Joshua Greene and Jonathan Haidt. How (and where) does moral judgment work? *Trends in cognitive sciences*, 6(12):517–523, 2002.

[115] Dagmar Gromann and Maria M. Hedblom. Body-mind-language: Multilingual knowledge extraction based on embodied cognition. In *AIC*, pages 20–33, 2017.

[116] Dagmar Gromann and Maria M. Hedblom. Kinesthetic mind reader: A method to identify image schemas in natural language. In *Proceedings of Advancements in Cogntivie Systems*, 2017.

[117] Tom Gruber. What is an ontology, 1993. Online: http://www-ksl.stanford.edu/kst/what-is-an-ontology.html.

[118] Nicola Guarino, Daniel Oberle, and Steffen Staab. What is an ontology? *Handbook on ontologies*, pages 1–17, 2009.

[119] Nicola Guarino and Christopher Welty. Evaluating ontological decisions with ontoclean. *Communications of the ACM*, 45(2):61–65, 2002.

[120] Giancarlo Guizzardi. Ontology, ontologies and the "I" of FAIR. *Data Intelligence*, 2(1-2):181–191, 2020.

[121] Giancarlo Guizzardi, Alessander Botti Benevides, Claudenir M Fonseca, Daniele Porello, João Paulo A Almeida, and Tiago Prince Sales. Ufo: Unified foundational ontology. *Applied Ontology*, 17(1):167–210, 2022.

[122] Qipeng Guo, Zhijing Jin, Ning Dai, Xipeng Qiu, Xiangyang Xue, David Wipf, and Zheng Zhang. P2: A plan-and-pretrain approach for knowledge graph-to-text generation: A plan-and-pretrain approach for knowledge graph-to-text generation. In *Proceedings of the 3rd International Workshop on Natural Language Generation from the Semantic Web (WebNLG+)*, pages 100–106, 2020.

[123] Jürgen Habermas. *On the Logic of the Social Sciences*. Cambridge: Polity Press, 1988. trans. by Shierry Weber Nicholsen and Jerry A. Stark.

[124] Christian Haerpfer, Ronald Inglehart, Alejandro Moreno, Christian Welzel, Kseniya Kizilova, Jaime Diez-Medrano, Marta Lagos, Pippa Norris, Eduard Ponarin, Bi Puranen, et al. World values survey: round seven–country-pooled datafile. *Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat*, 2020.

[125] Jonathan Haidt. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological review*, 108(4):814, 2001.

[126] Jonathan Haidt. *The moral emotions.* Oxford University Press, 2003.

[127] Jonathan Haidt. *The righteous mind: Why good people are divided by politics and religion*. Vintage, 2012.

[128] Jonathan Haidt and Matthew A. Hersh. Sexual morality: The cultures and emotions of conservatives and liberals 1. *Journal of Applied Social Psychology*, 31(1):191–221, 2001.

[129] Jonathan Haidt, Silvia Helena Koller, and Maria G. Dias. Affect, culture, and morality, or is it wrong to eat your dog? *Journal of personality and social psychology*, 65(4):613, 1993.

[130] Karl Hammar. *Content ontology design patterns: Qualities, methods, and tools*, volume 1879. Linköping University Electronic Press, 2017.

[131] Beate Hampe. Image schemas in cognitive linguistics: Introduction. *From perception to meaning: Image schemas in cognitive linguistics*, 29:1–14, 2005.

[132] Stephen Harding, David Phillips, and Michael Patrick Fogarty. *Contrasting values in Western Europe: Unity, diversity and change*. Macmillan Publishing Company, 1986.

[133] Janna Hastings, Werner Ceusters, Barry Smith, and Kevin Mulligan. The emotion ontology: enabling interdisciplinary research in the affective sciences. In *International and interdisciplinary conference on modeling and using context*, pages 119–123. Springer, 2011.

[134] Michael Hechter, Lynn Nadel, and Richard E. Michod. *The origin of values*. Aldine-Transaction, 1993.

[135] Maria M. Hedblom. *Image schemas and concept invention: cognitive, logical, and linguistic investigations*. Springer Nature, 2020.

[136] Maria M. Hedblom, Dagmar Gromann, and Oliver Kutz. In, Out and Through: Formalising some dynamic aspects of the image schema Containment. In Stefano Bistarelli, Martine Ceberio, Francesco Santini, and Eric Monfroy, editors, *Proceedings of the Knowledge Representation and Reasoning Track (KRR) at the Symposium of Applied Computing (SAC)*, pages 918–925, 2018.

[137] Maria M. Hedblom, Oliver Kutz, Till Mossakowski, and Fabian Neuhaus. Between contact and support: Introducing a logic for image schemas and directed movement. In *Conference of the Italian Association for Artificial Intelligence*, pages 256–268. Springer, 2017.

[138] Maria M. Hedblom, Oliver Kutz, and Fabian Neuhaus. Choosing the right path: image schema theory as a foundation for concept invention. *Journal of Artificial General Intelligence*, 6(1):21–54, 2015.

[139] Maria M. Hedblom, Oliver Kutz, Rafael Peñaloza, and Giancarlo Guizzardi. Image schema combinations and complex events. *KI-Künstliche Intelligenz*, 33(3):279–291, 2019.

[140] Patrick J. Henry and Christine Reyna. Value judgments: The impact of perceived value violations on american political attitudes. *Political Psychology*, 28(3):273–298, 2007.

[141] E. Tory Higgins. Value from hedonic experience and engagement. *Psychological review*, 113(3):439, 2006.

[142] Steven Hitlin and Jane Allyn Piliavin. Values: Reviving a dormant concept. *Annual review of sociology*, pages 359–393, 2004.

[143] Pascal Hitzler, Aldo Gangemi, and Krzysztof Janowicz. *Ontology engineering with ontology design patterns: foundations and applications*, volume 25. IOS Press, 2016.

[144] Geert Hofstede. The interaction between national and organizational value systems [1]. *Journal of management studies*, 22(4):347–357, 1985.

[145] Joe Hoover, Kate Johnson, Reihane Boghrati, Jesse Graham, and Morteza Dehghani. Moral framing and charitable donation: Integrating exploratory social media analyses and confirmatory experimentation. *Collabra: Psychology*, 4(1), 2018.

[146] Joe Hoover, Gwenyth Portillo-Wightman, Leigh Yeh, Shreya Havaldar, Aida Mostafazadeh Davani, Ying Lin, Brendan Kennedy, Mohammad Atari, Zahra Kamel, Madelyn Mendlen, et al. Moral foundations twitter corpus: A collection of 35k tweets annotated for moral sentiment. *Social Psychological and Personality Science*, 11(8):1057–1071, 2020.

[147] Joe Hoover, Gwenyth Portillo-Wightman, Leigh Yeh, Shreya Havaldar, Aida Mostafazadeh Davani, Ying Lin, Brendan Kennedy, Mohammad Atari, Zahra Kamel, Madelyn Mendlen, Gabriela Moreno, Christina Park, Tingyee E. Chang, Jenna Chin, Christian Leong, Jun Yen Leung, Arineh Mirinjian, and Morteza Dehghani. Moral foundations twitter corpus: A collection of 35k tweets annotated for moral sentiment. *Social Psychological and Personality Science*, 11(8):1057–1071, 2020.

[148] Frederic R. Hopp, Jacob T. Fisher, Devin Cornell, Richard Huskey, and René Weber. The extended moral foundations dictionary (emfd): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior research methods*, 53(1):232–246, 2021.

[149] Ioana Hulpuș, Jonathan Kobbe, Heiner Stuckenschmidt, and Graeme Hirst. Knowledge graphs meet moral values. In *Proceedings of the Ninth Joint Conference on Lexical and Computational Semantics*, pages 71–80, 2020.

[150] Jörn Hurtienne. Image schemas and design for intuitive use. 2011. Doctoral dissertation at Technische Universität Berlin.

[151] Jörn Hurtienne, Stephan Huber, and Cordula Baur. Supporting user interface design with image schemas: The iscat database as a research tool. 2022.

[152] Jörn Hurtienne and Johann Habakuk Israel. Image schemas and their metaphorical extensions: intuitive patterns for tangible interaction. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 127–134, 2007.

[153] Ronald Inglehart. Modernization, postmodernization and changing perceptions of risk. *International Review of Sociology*, 7(3):449–459, 1997.

[154] Carroll E. Izard. *The face of emotion.* Appleton-Century-Crofts, 1971.

[155] Olaf Jäkel. *Wie Metaphern Wissen schaffen: die kognitive Metapherntheorie und ihre Anwendung in Modell-Analysen der Diskursbereiche Geistestätigkeit, Wirtschaft, Wissenschaft und Religion*. Kovač, 2003.

[156] William James. The emotions. 1922.

[157] Ronnie Janoff-Bulman, Sana Sheikh, and Sebastian Hepp. Proscriptive versus prescriptive morality: two faces of moral regulation. *Journal of personality and social psychology*, 96(3):521, 2009.

[158] Hans Joas et al. *The genesis of values*. University of Chicago Press, 2000.

[159] Mark Johnson. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. The University of Chicago Press, Chicago and London, 1987.

[160] Roger Jowell, C. Roberts, Rory Fitzgerald, and Gillian Eva. the european social survey. *Measuring Attitudes Cross-Nationally: Lessons from the European Social Survey*, page 1, 2007.

[161] Cigdem Kagitcibasi and Ype H. Poortinga. Cross-cultural psychology. *Database*, 22:0221, 1989.

[162] Daniel Kahneman. *Thinking, fast and slow*. Farrar, Straus and Giroux, New York, 2011.

[163] Daniel Kahneman, Stewart Paul Slovic, Paul Slovic, and Amos Tversky. *Judgment under uncertainty: Heuristics and biases*. Cambridge university press, 1982.

[164] Gaetano Kanizsa. *Grammatica del vedere: saggi su percezione e gestalt*. Il mulino, 1980.

[165] Immanuel Kant. Critique of pure reason. 1781. *Modern Classical Philosophers, Cambridge, MA: Houghton Mifflin*, pages 370–456, 1908.

[166] Rishemjit Kaur and Kazutoshi Sasahara. Quantifying moral foundations from various topics on twitter conversations. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 2505–2512. IEEE, 2016.

[167] C Maria Keet et al. The use of foundational ontologies in ontology development: An empirical assessment. In *ESWC (1)*, pages 321–335, 2011.

[168] Brendan Kennedy, Mohammad Atari, Aida Mostafazadeh Davani, Joe Hoover, Ali Omrani, Jesse Graham, and Morteza Dehghani. Moral concerns are differentially observable in language. *Cognition*, 212:104696, 2021.

[169] John Maynard Keynes. The general theory of employment. *The quarterly journal of economics*, 51(2):209–223, 1937.

[170] Qaisar Khan and Huina Chua. An automated topics labeling framework using zero-shot text classification. *Journal of Engineering Science and Technology Special Issue on ACSAT*, December (2021) 46 - 59.

[171] R.W. Kilby. The study of human values. lamhan, 1993.

[172] Seonhoon Kim, Inho Kang, and Nojun Kwak. Semantic sentence matching with densely-connected recurrent and co-attentive information, 2018.

[173] Michael Kimmel. Culture regained: Situated and compound image schemas. *From perception to meaning: Image schemas in cognitive linguistics*, pages 285–311, 2005.

[174] Michael Kimmel. Analyzing image schemas in literature. *Cognitive Semiotics*, 5(1-2):159–188, 2009.

[175] K. Kipper-Schuler. VerbNet: A broad-coverage, comprehensive verb lexicon (Ph. D. thesis). *University of Pennsylvania, Philadelphia, PA*, 1005, 2006.

[176] Clyde Kluckhohn. Values and value-orientations in the theory of action. *Toward a General Theory of Action, op. cit*, pages 394–403, 1952.

[177] Clyde Kluckhohn. *Values and value-orientations in the theory of action: An exploration in definition and classification*. Harvard University Press, 2013.

[178] Jonathan Kobbe, Ines Rehbein, Ioana Hulpus, and Heiner Stuckenschmidt. Exploring morality in argumentation. Association for Computational Linguistics, ACL, 2020.

[179] Lawrence Kohlberg. *Essays on moral development/2 The psychology of moral development*. Harper & Row, 1984.

[180] Dennis L. Krebs. Morality: An evolutionary account. *Perspectives on psychological science*, 3(3):149–172, 2008.

[181] Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A Shamma, et al. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International journal of computer vision*, 123(1):32–73, 2017.

[182] Fred Kröger and Stephan Merz. *Temporal Logic and State Systems (Texts in Theoretical Computer Science. An EATCS Series)*. Springer Publishing Company, Incorporated, 1 edition, 2008.

[183] Thomas Kuhn. The structure of scientific revolutions. *International Encyclopedia of Unified Science*, 2(2), 1962.

[184] Werner Kuhn. An Image-Schematic Account of Spatial Categories. In Stephan Winter, Matt Duckham, Lars Kulik, and Ben Kuipers, editors, *Spatial Information Theory*, volume 4736 of *LNCS*, pages 152–168. Springer, 2007.

[185] Jouni-Matti Kuukkanen. Making sense of conceptual change. *History and theory*, 47(3):351–372, 2008.

[186] Haewoon Kwak, Jisun An, Elise Jing, and Yong-Yeol Ahn. Frameaxis: characterizing microframe bias and intensity with word embedding. *PeerJ Computer Science*, 7:e644, 2021.

[187] George Lakoff. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. The University of Chicago Press, 1987.

[188] George Lakoff. Conceptual metaphor. *Cognitive Linguistics: Basic Readings. Berlin*, pages 185–239, 2006.

[189] George Lakoff. *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago press, 2008.

[190] George Lakoff and Mark Johnson. *Metaphors we live by*. University of Chicago press, 1980.

[191] George Lakoff and Mark Johnson. *Philosophy in the flesh: The embodied mind and its challenge to western thought*, volume 640. Basic books New York, 1999.

[192] Ronald W. Langacker. *Foundations of cognitive grammar: Theoretical prerequisites*, volume 1. Stanford university press, 1987.

[193] Carl Georg Lange and Istar A. Haupt. The emotions. 1922.

[194] Daniel K. Lapsley and Darcia Narvaez. A social–cognitive approach to the moral personality. In *Moral development, self, and identity*, pages 201–224. Psychology Press, 2004.

[195] Richard S. Lazarus. Cognition and motivation in emotion. *American psychologist*, 46(4):352, 1991.

[196] Beth Levin. *English verb classes and alternations: A preliminary investigation*. University of Chicago press, 1993.

[197] Manling Li, Ying Lin, Joseph Hoover, Spencer Whitehead, Clare Voss, Morteza Dehghani, and Heng Ji. Multilingual entity, relation, event and human value extraction. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 110–115, 2019.

[198] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. Dailydialog: A manually labelled multi-turn dialogue dataset. *arXiv:1710.03957*, 2017.

[199] Antonio Lieto, Luigi Asprino, Manuel Striani, Stefano De Giorgis, Rossana Damiano, and Aldo Gangemi. D6.3 Prototype of Knowledge based Support Deliverable 6.3, H2020 SPICE Project, 2022.

[200] Ying Lin, Joe Hoover, Gwenyth Portillo-Wightman, Christina Park, Morteza Dehghani, and Heng Ji. Acquiring background knowledge to improve moral value prediction. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 552–559. IEEE, 2018.

[201] Nancy F. Link, Shawn E. Scherer, and P. Niall Byrne. Moral judgment and moral conduct in the psychopath. *Canadian Psychiatric Association Journal*, 22(7):341–346, 1977.

[202] Enrico Liscio, Alin Dondera, Andrei Geadau, Catholijn Jonker, and Pradeep Murukannaiah. Cross-domain classification of moral values. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 2727–2745, 2022.

[203] Enrico Liscio, M.T. van der Meer, Luciano Cavalcante Siebert, Catholijn M. Jonker, Niek Mouter, and Pradeep K. Murukannaiah. Axies: Identifying and evaluating context-specific values. In *Proceedings of the 20th international conference on Autonomous Agents and MultiAgent Systems*, pages 799–808. International Foundation for Autonomous Agents and Multiagent Systems, 2021.

[204] Hui Liu, Yinghui Huang, Zichao Wang, Kai Liu, Xiangen Hu, and Weijun Wang. Personality or value: A comparative study of psychographic segmentation based on an online review enhanced recommender system. *Applied Sciences*, 9(10):1992, 2019.

[205] George Loewenstein. Emotions in economic theory and economic behavior. *American Economic Review*, 90(2):426–432, 2000.

[206] Juan Miguel López Gil, Roberto García González, Rosa María Gil Iranzo, and César A Collazos Ordóñez. Emotionsonto: an ontology for developing affective applications. *Journal Of Universal Computer Science, 2014, vol. 13, num. 20, p. 1813-1828*, 2014.

[207] Catherine Lutz and Geoffrey M. White. The anthropology of emotions. *Annual review of anthropology*, pages 405–436, 1986.

[208] Alasdair MacIntyre. *After virtue*. A&C Black, 2013.

[209] Salvatore R. Maddi. *Personality theories: A comparative analysis*. Thomson Brooks/Cole Publishing Co, 1996.

[210] Manamohan Mahapatra and Lawrence Park. Divinity and the “Big Three" Explanations of Suffering. *Morality and Health*, page 119, 1997.

[211] Jean M. Mandler. *How to build a baby: Conceptual primitives.*, volume 99. American Psychological Association, 1992.

[212] Jean M. Mandler. *The foundations of mind: Origins of conceptual thought*. Oxford University Press, 2004.

[213] Jean M. Mandler. Thought before language. *Trends in cognitive sciences*, 8(11):508–513, 2004.

[214] Jean M. Mandler and Cristóbal Pagán Cánovas. On defining image schemas. *Language and Cognition*, pages 1–23, 2014.

[215] Jean M. Mandler and Cristóbal Pagán Cánovas. On defining image schemas. *Language and cognition*, 6(4):510–532, 2014.

[216] Mitchell Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. Building a large annotated corpus of english: The penn treebank. 1993.

[217] Margaret Mooney Marini. Social values and norms. *Encyclopedia of sociology*, 4:2828–2840, 2000.

[218] Alfred Marshall. *Principles of economics*. 2009. Unabridged Eighth Edition.

[219] Sebastiaan Mathôt, Daniel Schreij, and Jan Theeuwes. Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods*, 44(2):314–324, 2012.

[220] Albert Mehrabian. Comparison of the pad and panas as models for describing emotions and for differentiating anxiety from depression. *Journal of psychopathology and behavioral assessment*, 19(4):331–357, 1997.

[221] Igor Melnyk, Pierre Dognin, and Payel Das. Knowledge graph generation from text. *arXiv:2211.10511*, 2022.

[222] Maria Rosaria Stufano Melone, Stefano Borgo, and Oliver Kutz. Image schemas and ontology in architectural rules. In *The Eighth Joint Ontology Workshops (JOWO'22), CEUR volume 3249*, 2022.

[223] John Mikhail. Universal moral grammar: Theory, evidence and the future. *Trends in cognitive sciences*, 11(4):143–152, 2007.

[224] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv:1301.3781*, 2013.

[225] George A. Miller. *WordNet: An electronic lexical database*. MIT press, 1998.

[226] M. Minsky, A. Collins, and E.E. Smith. A framework for representing knowledge marvin minsky. In *Cognitive Science*. Morgan-Kaufmann, 1992.

[227] Saif M. Mohammad and Felipe Bravo-Marquez. Wassa-2017 shared task on emotion intensity. *arXiv:1708.03700*, 2017.

[228] Negar Mokhberian, Andrés Abeliuk, Patrick Cummings, and Kristina Lerman. Moral framing and ideological bias of news. In *International Conference on Social Informatics*, pages 206–219. Springer, 2020.

[229] Till Mossakowski, Mihai Codescu, Fabian Neuhaus, and Oliver Kutz. The distributed ontology, modeling and specification language–DOL. In *The Road to Universal Logic*, pages 489–520. Springer, 2015.

[230] Roberto Navigli and Simone Paolo Ponzetto. Babelnet: Building a very large multilingual semantic network. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 216–225, 2010.

[231] Fabian Neuhaus. What is an ontology? *arXiv preprint arXiv:1810.09171*, 2018.

[232] Pegah Nokhiz and Fengjun Li. Understanding rating behavior based on moral foundations: The case of yelp reviews. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 3938–3945. IEEE, 2017.

[233] Daniel Nowack and Sophia Schoderer. The role of values for social cohesion: Theoretical explication and empirical exploration. *German Development Institute, Discussion Paper*, 6, 2020.

[234] Larry P. Nucci and Elliot Turiel. Social interactions and the development of social concepts in preschool children. *Child development*, pages 400–407, 1978.

[235] Martha Nussbaum. Nature, Function and Capability. In *Oxford Studies in Ancient Philosophy: Supplementary Volume*. Oxford University Press, 1987.

[236] Andrea Giovanni Nuzzolese, Aldo Gangemi, and Valentina Presutti. Gathering lexical linked data and knowledge patterns from framenet. In *Proceedings of the sixth international conference on Knowledge capture*, pages 41–48. ACM, 2011.

[237] Todd Oakley. Image schemas. pages 214–235. Oxford University Press Oxford, 2007.

[238] Andrew Ortony. Are emotion metaphors conceptual or lexical? *Cognition and Emotion*, 2(2):95–104, 1988.

[239] Andrew Ortony, Gerald L. Clore, and Allan Collins. *The cognitive structure of emotions*. Cambridge university press, 2022.

[240] Anna Papafragou, Christine Massey, and Lila Gleitman. When English proposes what Greek presupposes: The cross-linguistic encoding of motion events. *Cognition*, 98(3):B75–B87, 2006.

[241] Talcott Parsons. *The Structure of Social Action*, volume 491. McGraw-Hill Book Company, Incorporated, 1937.

[242] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[243] Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H. Miller, and Sebastian Riedel. Language models as knowledge bases? *arXiv:1909.01066*, 2019.

[244] Mary L. Phillips, Wayne C. Drevets, Scott L. Rauch, and Richard Lane. Neurobiology of emotion perception I: The neural basis of normal emotion perception. *Biological psychiatry*, 54(5):504–514, 2003.

[245] Jean Piaget. *Cognitive Development in Children*. Springer, 1976.

[246] Stephen M. Pittel and Gerald A. Mendelsohn. Measurement of moral values: A review and critique. *Psychological Bulletin*, 66(1):22, 1966.

[247] Robert Plutchik. A psychoevolutionary theory of emotions, 1982.

[248] Robert Plutchik. *Emotions and life: Perspectives from psychology, biology, and evolution.* American Psychological Association, 2003.

[249] Robert Plutchik and Albert F. Ax. A critique of determinants of emotional state by schachter and singer (1962). *Psychophysiology*, 4(1):79–82, 1967.

[250] Robert Plutchik and Henry Kellerman. *Emotion, theory, research, and experience*, volume 3. Academic press, 1980.

[251] Louis P. Pojman. *Ethical theory: Classical and contemporary readings*. Wadsworth, 1995.

[252] Vladimir Ponizovskiy, Murat Ardag, Lusine Grigoryan, Ryan Boyd, Henrik Dobewall, and Peter Holtz. Development and validation of the personal values dictionary: A theory–driven tool for investigating references to basic human values in text. *European Journal of Personality*, 34(5):885–902, 2020.

[253] David Popenoe. American family decline, 1960-1990: A review and appraisal. *Journal of Marriage and the Family*, pages 527–542, 1993.

[254] Valentina Presutti and Aldo Gangemi. Content ontology design patterns as practical building blocks for web ontologies. In *International Conference on Conceptual Modeling*, pages 128–141. Springer, 2008.

[255] Juan Antonio Prieto Velasco and Maribel Tercedor Sánchez. The embodied nature of medical concepts: image schemas and language for PAIN. *Cognitive processing*, January 2014.

[256] J. Hunter Priniski, Negar Mokhberian, Bahareh Harandizadeh, Fred Morstatter, Kristina Lerman, Hongjing Lu, and P. Jeffrey Brantingham. Mapping moral valence of tweets following the killing of George Floyd. *arXiv:2104.09578*, 2021.

[257] Jesse Prinz. Which emotions are basic. *Emotion, evolution, and rationality*, 69:88, 2004.

[258] Jesse J. Prinz and Shaun Nichols. Moral emotions. *The Moral Psychology Handbook*, 2010. Oxford University Press.

[259] James Pustejovsky. Type theory and lexical decomposition. In *Advances in generative lexicon theory*, pages 9–38. Springer, 2013.

[260] James Pustejovsky and Pierrette Bouillon. Aspectual coercion and logical polysemy. *Journal of Semantics*, 12(2):133–162, 1995.

[261] James Pustejovsky, Adam Meyers, Martha Palmer, and Massimo Poesio. Merging propbank, nombank, timebank, penn discourse treebank and coreference. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pages 5–12, 2005.

[262] James Pustejovsky and Jessica L Moszkowicz. The qualitative spatial dynamics of motion in language. *Spatial Cognition & Computation*, 11(1):15–44, 2011.

[263] Boele de Ed Raad and Marco Ed Perugini. *Big five factor assessment: Introduction.* Hogrefe & Huber Publishers, 2002.

[264] David A. Randell, Zhan Cui, and Anthony G. Cohn. A spatial logic based on regions and connection. In *Proc. 3rd Int. Conf. on Knowledge Representation and Reasoning*, 1992.

[265] Revanth Gangi Reddy, Sai Chinthakindi, Zhenhailong Wang, Yi R. Fung, Kathryn S. Conger, Ahmed S. Elsayed, Martha Palmer, and Heng Ji. Newsclaims: A new benchmark for claim detection from news with background knowledge. *arXiv:2112.08544*, 2021.

[266] Mark Reynolds. The complexity of temporal logic over the reals. *Annals of Pure and Applied Logic*, 161(8), 2010.

[267] Rezvaneh Rezapour, Saumil H. Shah, and Jana Diesner. Enhancing the measurement of social effects by capturing morality. In *Proceedings of the tenth workshop on computational approaches to subjectivity, sentiment and social media analysis*, pages 35–45, 2019.

[268] Mariela Rico, María Laura Caliusco, Omar Chiotti, and María Rosa Galli. Ontoqualitas: A framework for ontology quality assessment in information interchanges between heterogeneous systems. *Computers in Industry*, 65(9):1291–1300, 2014.

[269] Adam Roberts, Colin Raffel, and Noam Shazeer. How much knowledge can you pack into the parameters of a language model? *arXiv:2002.08910*, 2020.

[270] John P. Robinson, Phillip R. Shaver, and Lawrence S. Wrightsman. Criteria for scale selection and evaluation. *Measures of personality and social psychological attitudes*, 1:1–16, 1991.

[271] Michael D. Robinson, Edward R. Watkins, and Eddie Harmon-Jones. *Handbook of cognition and emotion*. Guilford Press, 2013.

[272] Meg J. Rohan. A rose by any name? the values construct. *Personality and social psychology review*, 4(3):255–277, 2000.

[273] Milton Rokeach. A theory of organization and change within value-attitude systems. *Journal of social issues*, 1968.

[274] Milton Rokeach. *The nature of human values*. Free press, 1973.

[275] Ira J. Roseman, Ann Aliki Antoniou, and E. Paul. Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory. *Cognition and Emotion*, (3):241–277, 1996. Routledge.

[276] Sara Rosenthal, Noura Farra, and Preslav Nakov. Semeval-2017 task 4: Sentiment analysis in twitter. *arXiv:1912.00741*, 2019.

[277] Paul Rozin, Laura Lowery, Sumio Imada, and Jonathan Haidt. The CAD triad hypothesis: a mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of personality and social psychology*, 76(4):574, 1999.

[278] James A. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.

[279] David Ed Sander and Klaus R. Scherer. *The Oxford companion to emotion and the affective sciences*. Oxford University Press, 2009.

[280] Francisco Santibáñez. The object image-schema and other dependent schemas. *Atlantis*, pages 183–201, 2002.

[281] Francisco Santibáñez. The object image-schema and other dependent schemas. *Atlantis*, 24(2):183–201, 2002.

[282] Stanley Schachter. The interaction of cognitive and physiological determinants of emotional state. In *Advances in experimental social psychology*, volume 1, pages 49–80. Elsevier, 1964.

[283] Roger C. Schank and Robert P. Abelson. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Psychology Press, 2013.

[284] Klaus R. Scherer and Harald G. Wallbott. Evidence for universality and cultural variation of differential emotion response patterning. *Journal of personality and social psychology*, 66(2):310, 1994.

[285] Marco Schorlemmer, Roberto Confalonieri, and Enric Plaza. The Yoneda Path to the Buddhist Monk Blend. In *Proceedings of the Joint Ontology Workshops 2016 Episode 2*, 2016.

[286] Karin Kipper Schuler. *VerbNet: A broad-coverage, comprehensive verb lexicon*. University of Pennsylvania, 2005.

[287] Shalom Schwartz. A theory of cultural value orientations: Explication and applications. *Comparative sociology*, 5(2-3):137–182, 2006.

[288] Shalom H. Schwartz. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology*, volume 25, pages 1–65. Elsevier, 1992.

[289] Shalom H. Schwartz. Schwartz value survey. *Journal of Cross-Cultural Psychology*, 2005.

[290] Shalom H. Schwartz. Basic human values: Theory, methods, and application. *Basic Human Values*, pages 1000–1023, 2007.

[291] Shalom H. Schwartz. An overview of the schwartz theory of basic values. *Online readings in Psychology and Culture*, 2(1):2307–0919, 2012.

[292] Shalom H. Schwartz and Anat Bardi. Value hierarchies across cultures: Taking a similarities perspective. *Journal of cross-cultural Psychology*, 32(3):268–290, 2001.

[293] Shalom H. Schwartz and Wolfgang Bilsky. Toward a universal psychological structure of human values. *Journal of personality and social psychology*, 53(3):550, 1987.

[294] Shalom H. Schwartz and Wolfgang Bilsky. Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of personality and social psychology*, 58(5):878, 1990.

[295] Shalom H Schwartz and Klaus Boehnke. Evaluating the structure of human values with confirmatory factor analysis. *Journal of research in personality*, 38(3):230–255, 2004.

[296] Shalom H. Schwartz, Jan Cieciuch, Michele Vecchione, Eldad Davidov, Ronald Fischer, Constanze Beierlein, Alice Ramos, Markku Verkasalo, Jan-Erik Lönnqvist, Kursad Demirutku, et al. Refining the theory of basic individual values. *Journal of personality and social psychology*, 103(4):663, 2012.

[297] Shalom H. Schwartz, Gila Melech, Arielle Lehmann, Steven Burgess, Mari Harris, and Vicki Owens. Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of cross-cultural psychology*, 32(5):519–542, 2001.

[298] William A. Scott. Empirical assessment of values and ideologies. *American Sociological Review*, pages 299–310, 1959.

[299] Amartya Sen. *Choice, welfare and measurement*. Harvard University Press, 1997.

[300] Lawrence Shapiro. *Embodied cognition*. Routledge, 2010.

[301] Phillip Shaver, Judith Schwartz, Donald Kirson, and Cary O'connor. Emotion knowledge: further exploration of a prototype approach. *Journal of personality and social psychology*, 52(6):1061, 1987.

[302] Taylor Shin, Yasaman Razeghi, Robert L. Logan I.V., Eric Wallace, and Sameer Singh. Autoprompt: Eliciting knowledge from language models with automatically generated prompts. *arXiv:2010.15980*, 2020.

[303] Jordi Simón. The portrait values questionnaire: A bibliographic and bibliometric review of the instrument. *Aloma: Revista de Psicologia, Ciències de l'Educació i de l'Esport*, 35(1):39–50, 2017.

[304] Linda J. Skitka, Christopher W. Bauman, and Edward G. Sargis. Moral conviction: Another contributor to attitude strength or something more? *Journal of personality and social psychology*, 88(6):895, 2005.

[305] Adam Smith. *The wealth of nations [1776]*, volume 11937. na, 1937.

[306] Barry Smith, Anand Kumar, and Thomas Bittner. Basic formal ontology for bioinformatics. 2005.

[307] M. Brewster Smith. Personal values in the study of lives. 1963.

[308] Robyn Speer, Joshua Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the AAAI conference on Artificial Intelligence*, volume 31, 2017.

[309] Robert St. Amant, Clayton T. Morrison, Yu-Han Chang, Paul R. Cohen, and Carole Beal. An image schema language. In *Proc. of the 7th Int. Conf. on Cognitive Modeling (ICCM)*, pages 292–297, 2006.

[310] Jacopo Staiano and Marco Guerini. Depechemood: a lexicon for emotion analysis from crowd-annotated news. *arXiv:1405.1605*, 2014.

[311] Mark Steedman. Formalizing affordance. In *Proc. of the 24th Annual Meeting of the Cognitive Science Society*, pages 834–839. Routledge, 2002.

[312] Richard I. Steim and Carol J. Nemeroff. Moral overtones of food: Judgments of others based on what they eat. *Personality and Social Psychology Bulletin*, 21(5):480–490, 1995.

[313] Carlo Strapparava, Alessandro Valitutti, et al. WordNet Affect: an affective extension of wordnet. In *Lrec*, volume 4, page 40. Lisbon, Portugal, 2004.

[314] Nina Strohminger and Victor Kumar. *The moral psychology of disgust*. Rowman & Littlefield, 2018.

[315] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Mariano Fernández-López. The neon methodology for ontology engineering. In *Ontology engineering in a networked world*, pages 9–34. Springer, 2012.

[316] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Mariano Fernandez-Lopez. The neon methodology framework: A scenario-based methodology for ontology development. *Applied ontology*, 10(2):107–145, 2015.

[317] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. YAGO: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*, pages 697–706, 2007.

[318] Swabha Swayamdipta, Sam Thomson, Chris Dyer, and Noah A. Smith. Frame-Semantic Parsing with Softmax-Margin Segmental RNNs and a Syntactic Scaffold. *arXiv:1706.09528*, 2017.

[319] Leonard Talmy. How language structures space. In *Spatial Orientation*, pages 225–282. Springer, 1983.

[320] Leonard Talmy. Force dynamics in language and cognition. *Cognitive science*, 12(1):49–100, 1988.

[321] Leonard Talmy. The fundamental system of spatial schemas in language. In Beate Hampe and Joseph E Grady, editors, *From perception to meaning: Image schemas in cognitive linguistics*, volume 29 of *Cognitive Linguistics Research*, pages 199–234. Walter de Gruyter, 2005.

[322] Sang-Sang Tan and Jin-Cheon Na. Positional Attention-based Frame Identification with BERT: A Deep Learning Approach to Target Disambiguation and Semantic Frame Selection, 2019.

[323] Mark B. Tappan. Mediated moralities: Sociocultural approaches to moral development. In *Handbook of moral development*, pages 369–392. Psychology Press, 2006.

[324] Linnet Taylor, Luciano Floridi, and Bart Van der Sloot. *Group privacy: New challenges of data technologies*, volume 126. Springer, 2016.

[325] Livia Teernstra, Peter van der Putten, Liesbeth Noordegraaf-Eelens, and Fons Verbeek. The morality machine: Tracking moral values in tweets. In *International Symposium on Intelligent Data Analysis*, pages 26–37. Springer, 2016.

[326] Marco Tettamanti, Giovanni Buccino, Maria Cristina Saccuman, Vittorio Gallese, Massimo Danna, Paola Scifo, Ferruccio Fazio, Giacomo Rizzolatti, Stefano F Cappa, and Daniela Perani. Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of cognitive neuroscience*, 17(2):273–281, 2005.

[327] Jackson Trager, Alireza S. Ziabari, Aida Mostafazadeh Davani, Preni Golazazian, Farzan Karimi-Malekabadi, Ali Omrani, Zhihe Li, Brendan Kennedy, Nils Karl Reimer, Melissa Reyes, et al. The Moral Foundations Reddit Corpus. *arXiv:2208.05545*, 2022.

[328] Harry Charalambos Triandis. *Handbook of cross-cultural psychology*, volume 4. Allyn & Bacon, 1980.

[329] Elliot Turiel. Thought about actions in social domains: Morality, social conventions, and social interactions. *Cognitive development*, 23(1):136–154, 2008.

[330] Jonathan H. Turner and Jan E. Stets. *The sociology of emotions*. Cambridge University Press, 2005.

[331] Michael Uschold. Demystifying owl for the enterprise. *Synthesis Lectures on Semantic Web: Theory and Technology*, 8(1):i–237, 2018.

[332] Baastian C. Van Fraassen. Values and the heart's command. *The Journal of Philosophy*, 70(1):5–19, 1973.

[333] F.J. Varela, E. Thompson, and E. Rosch. The Embodied Mind: Cognitive Science and Human Experience. *Psychological Medicine*, 23(2):529–532, 1993.

[334] Christin-Melanie Vauclair. Measuring cultural values at the individual-level: considering morality in cross-cultural value research. *RAM. Revista de Administração Mackenzie*, 10:60–83, 2009.

[335] Christin-Melanie Vauclair. *What is right? What is wrong? And does the answer tell something about culture? - An investigation into culture and morality using the values concept*. Victoria University of Wellington, 2010.

[336] Ian Verstegen. Gestalt psychology in italy. *Journal of the History of the Behavioral Sciences*, 36(1):31–42, 2000.

[337] Patrizia Violi. Embodiment at the crossroads between cognition and semiosis. *Recherches en communication*, 19:199–217, 2003.

[338] Lennart Wachowiak. Semi-automatic extraction of image schemas from natural language. In *Proceedings of the MEi:CogSci Conference 2020*, page 105. Comenius University, Bratislava, 2020.

[339] Lennart Wachowiak and Dagmar Gromann. Systematic Analysis of Image Schemas in Natural Language through Explainable Multilingual Neural Language Processing. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 5571–5581.

[340] Lisa Walton and Michael Worboys. An Algebraic Approach to Image Schemas for Geographic Space. In *Spatial Information Theory*, volume 5756 of *Lecture Notes in Computer Science*, pages 357–370. Springer, 2009.

[341] Bo Wang, Maria Liakata, Arkaitz Zubiaga, Rob Procter, and Eric Jensen. Smiles: Twitter emotion classification using domain. In *SAAIP@ IJCAI*, 2016.

[342] Chenguang Wang, Xiao Liu, and Dawn Song. Language models are open knowledge graphs. *arXiv:2010.11967*, 2020.

[343] Shenghui Wang, Stefan Schlobach, and Michel Klein. Concept drift and how to identify it. *Journal of Web Semantics*, 9(3):247–265, 2011.

[344] Max Weber. The theory of economic and social organization. *Translation by A.M. Henderson and Talcott Parsons*, 1947.

[345] Nico Van De Weghe, Anthony G. Cohn, Guy De Tré, and Philippe De Maeyer. A qualitative trajectory calculus as a basis for representing moving objects in geographical information systems. *Control and cybernetics*, 35(1):97–119, 2006.

[346] Mark Wick and Bernard Vatant. The geonames geographical database. *Available from World Wide Web: http://geonames. org*, 2012.

[347] Terry Winograd. Frame representations and the declarative/procedural controversy. In *Representation and understanding*, pages 185–210. Elsevier, 1975.

[348] Jing Yi Xie, Renato Ferreira Pinto Jr., Graeme Hirst, and Yang Xu. Text-based inference of moral sentiment change. *arXiv:2001.07209*, 2020.

[349] Wenpeng Yin, Jamaal Hay, and Dan Roth. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach, 2019.

[350] Hong Zhu, Dongmei Liu, Ian Bayley, Arantza Aldea, Yunfei Yang, and Ying Chen. Quality model and metrics of ontology for semantic descriptions of web services. *Tsinghua science and technology*, 22(3):254–272, 2017.