"The time for action has arrived": Extending the IS Catalogue leveraging Large Language Models

Stefano **De Giorgis**[*,†1], Guendalina **Righetti**[*,†2]

[1]*Institute for Cognitive Sciences and Technologies - National Research Council (ISTC-CNR), Italy*

[2]*Department of Philosophy, Classics, History of Art and Ideas, University of Oslo, Blindernveien 31 Georg Morgenstiernes hus 0313 Oslo*

Abstract

Image Schema research has long been hampered by the scarcity of annotated data, limiting advancements in the field. This paper presents a novel approach to overcoming this challenge by leveraging Large Language Models (LLMs) to extend the IS Catalogue. We systematically tested various LLMs to identify the most effective model for this task, ultimately selecting Claude 3.5 Sonnet. We asked the model to extend the IS catalogue in two ways: first, by expanding the annotations associated with the sentences in the catalogue to include multiple IS; second, by generating new literal sentences to be added to the catalogue. To evaluate the model we conducted several analyses, including accuracy ratings with the original annotations. Our approach demonstrated remarkable efficacy, with the chosen model successfully retrieving the original annotation in 81% of cases when considering the entire set of image schemas extracted in the profile. This approach enables rapid processing and annotation of large text volumes while maintaining high accuracy and consistency. Partial evaluation by domain experts has found the enriched IS Catalogue to be sound and plausible, suggesting that LLM-assisted extension can produce high-quality synthetic data aligned with expert knowledge. Our method offers a promising solution to the data scarcity problem in IS research, potentially accelerating advancements in the field.

1. Introduction

Image schemas (IS) are foundational conceptual structures within the paradigm of embodied cognition. These schemas encapsulate sensorimotor experiences and play a crucial role in shaping abstract cognition, including commonsense reasoning and the semantic underpinnings of natural language (see e.g. Mandler and Hampe [1, 2]). As internally structured gestalts [3], image schemas are composed of spatial primitives (SP) that coalesce into unified wholes of meaning, thereby forming more complex schematic structures [4, 1, 5].

The current main IS repository is the Image Schema Catalogue [6, 7]. While valuable, there are some problems to it: (i) each sentence is annotated with only one IS, and (ii) the list of IS used is not comprehensive, due to the open debate about the final full list. As a result, a single annotation often oversimplifies the rich, multi-layered nature of image schematic conceptual structures embedded in everyday language as well as in conceptual metaphors. Consider the sentence "Sally found an idea in the book," which is annotated solely with the OBJECT schema in the catalogue. This annotation, while not incorrect, fails to capture the full conceptual richness of the expression. The presence of the preposition 'in' clearly activates the CONTAINMENT schema, suggesting that ideas are conceptualized as entities, that can be contained within physical or non-physical objects (in this case, a book). This example underscores a critical issue: the overlapping and often inseparable nature of image schemas in natural language. The OBJECT schema (applied to both 'idea' and 'book') and the CONTAINMENT schema are not merely co-present but fundamentally intertwined in conveying the sentence's meaning.

Another level of problematicity (iii) is given by the scarcity of data. More specifically, in the Image Schema Catalogue, the sentences gathered exemplify mostly metaphoric usage of image schemas (see

---

above the "Sally found an idea in the book" example). While this focus is valuable when the catalogue is used for research in conceptual metaphor or blending, incorporating examples of more concrete applications would be advantageous for studies involving practical scenarios, such as applying image schema research to robotics and similar fields. In such cases, providing concrete examples could also increase the pedagogical value of the catalogue.

In this work, we tackle problems (i) and (iii), namely the limited IS Catalogue annotation and the scarcity of available data, and we do so by exploiting synthetic data generation via the usage of Generative AI in the form of Large Language Models (LLM).

We propose the following pipeline. 1. We compare several state-of-the-art LLMs on a classification task to test their ability to identify image schematic knowledge from natural language; 2. Once identified the best model, we pass each sentence of the IS Catalogue to the LLM, asking it to indicate the image schema(s) evoked by the sentence, ranking them in descending order from the most relevant to the least; 3. We ask the model to reproduce, for each metaphoric sentence contained in the catalogue, a corresponding literal sentence replicating the same image-schematic pattern found in the metaphorical one (with results like this: "The time for action has arrived" → "The train has arrived to the station"). We conducted several analyses to shed light on Claude's actual capabilities of mastering the identification and reuse of image schematic content in natural language sentences, including annotation matching, Is distribution analysis, co-occurrence and confusion matrices analyses. Overall, we achieve an 81% precision comparing the newly produced annotation with the original one, and we manually validate (part of) the extended IS Catalogue with the help of domain experts.

The paper is organised as follows: Section 2 provide some useful references to image schema and IS Catalogue usage; Section 3 details the methodology, prompting strategy and technical details of the approach; in Section 4 we show some quantitative analysis of the IS Catalogue extension and we reflect on interesting findings of these experiments. Finally Section 5 concludes the paper and opens to possible future works.

2. Background and Related Work

Image schemas, firstly introduced by Lakoff and Johnson [3, 8, 9] are now recognised as sensorimotor cognitive patterns shaping our way of conceive the world and establish semantic relations, based on our bodily perception [10, 11, 12, 13, 14]. Recent significant efforts to investigate image schemas and their compositional nature include the development of Image Schema Logic ISL^{FOL} [15] their capabilities in conceptual blending [16, 17], and the ImageSchemaNet ontology [18]. These frameworks provide robust tools for analyzing phenomena such as conceptual blending and cognitive metaphors, as well as image schematic analysis of complex events [19]. Image schemas can be represented also via a diagrammatic image schema language [20].

While corpus-based studies [21, 22] and machine learning approaches [23, 24, 25] have explored the presence of image schemas in natural language, the complexity of image schema annotation in natural language still presents a significant challenge in the field of cognitive linguistics and computational semantics.

We refer to IS Catalogue in its version enriched with the MetaNet source and target domains alignment[1]. The dataset includes linguistic examples taken from several resources (MetaNet, Lakoff and Johnson works, Dodge [12], etc.) and for each example it individuates the corresponding conceptual metaphor, source and target domain, the "sensorimotor" source domain - namely the embodied grounding of the sentence, which sometimes is a single spatial primitive, sometimes a full image schema -

---

[1]Available here: https://github.com/dgromann/ImageSchemaRepository

and a dedicated column to the image schema evoked by the sentence. The IS Catalogue has been used previously in several works, with different purposes, for example in ODIN [26] for identifying image schematic grounding in Ontology Design Patterns; as a support tool for User Interface Design tasks [7], and in linguistic tasks to train a supervised classifier to classify natural language expressions in order to detect image schemas from multilingual inputs [25].

In this context, we employ the term "'image schema profile" as defined by [27] and [28]. This concept refers to the collective set of activated image schemas associated with a particular entity, sentence, situation, or event, providing a comprehensive framework for analyzing the schematic underpinnings of linguistic expressions. Specifically, we refer to the set of IS annotated by Claude for a sentence as "IS Profile" of that sentence. Note that, while in original definition of "IS Profile", the order of IS is not relevant, in our case it is, since, as detailed in Section 3, we instructed the LLM to order the annotations listing in descending order IS from the most relevant (the one fitting the most the analysed sentence) to the least relevant.

3. Methodology

In this section we provide details about the pipeline adopted to perform the LLM annotation, and dataset extension. The four main steps are: 1. choice of the best LLM model in terms of competence over the image schematic domain; 2. classification task, passing each sentence of the IS Catalogue for multi-label annotation; 3. Generation of literal twin sentences to the metaphorical original ones and 4. evaluation of the results.

Choice of the Model We conducted a comparative analysis of three leading language models: Anthropic's Claude 3.5 Sonnet, OpenAI's GPT-4o, and Google's Gemini. This evaluation aimed to assess their capability in identifying image schemas with minimal context and instruction. We employed a bulk, zero-shot approach, presenting each model with the simple prompt: "Which image schemas can you identify in the following sentence: What did you have in mind?". We prompted 5 different sentences taken from the catalogue and annotated with different image schemas. This method allowed us to gauge the models' innate understanding and retrieval of image schema concepts without additional training or context. We illustrate the performance of the models by discussing the example above, as the other cases were similar. Full answers are available on the dedicated GitHub repository: https://github.com/StenDoipanni/ISD8.

Gemini demonstrated the weakest performance, identifying only two schemas, one of which (Container) is more accurately classified as a spatial primitive rather than a full image schema, while the other (Possession) appeared to be a hallucination not typically included in standard image schema listings. GPT-4o showed similar results in the context of this specific example (while, in other cases, performed generally better than Gemini by providing more accurate results and providing credible justifications). However, Claude 3.5 Sonnet emerged as the top performer in this task. Not only did it correctly identify the highest number of plausible image schemas, but it also presented them in the most conventional format, selecting often the correct naming and using all-caps notation, which is standard in the field. For this reason we selected Claude 3.5 Sonnet as the primary model for our classification and generation tasks.

Classification Task Our approach leveraged the Claude 3.5 Sonnet model via the Anthropic API to annotate sentences with image schemas. We developed a Python script that processes sentences from the IS Catalogue in the form of a CSV file, sending each sentence to the Claude model for analysis. We kept all the original sentences included in the catalogue, including both English and German sentences. The model was prompted to perform the annotation task in the following way: (i) annotate the sentence with relevant image schemas from a predefined list, taken from IS Catalogue, and (ii) ordering them by relevance. The prompt adopted a few shot technique, providing three examples, presented in pseudo-json syntax as key-value pairs, annotated respectively with three, two, and four IS.

The script incorporates error handling and retry logic to manage potential API issues, with a maximum of 5 retries and a 5-second delay between attempts. The results, including the original sentence and image schema annotations were output in JSON format. To ensure reproducibility, the full prompt used for the Claude model is available in our GitHub repository.

Generation Task   Parallel to sentence classification the second part of the prompt was refined to generate a literal transpositions of metaphorical expressions, asking for more literal sentence preserving the same image schematic pattern. Again, the prompting technique adopted is a few shot, with three examples in pseudo-json syntax, presented as key-value pairs. To provide an example: *Metaphorical Example: "Our agenda is packed with events." Literal Sentence: "The bag is packed with clothes."*

Evaluation   We performed a number of quantitative analyses to assess Claude's ability to identify and reproduce image-schematic patterns, including (i) accuracy ratings, (ii) IS distribution analysis, (iii) co-occurrence and (iv) confusion matrices. Each analysis is diffusely described in the following sections. The primary evaluation focuses on the accuracy of the proposed annotations. Specifically, we assess (i) whether the image schema annotated in the catalogue matches the one identified by the LLM, and (ii) the position of the correct schema within the LLM's proposed ranking of preferences. Given the metaphorical and complex nature of the sentences in the catalogue, we consider the presence of the human-annotated IS among the profiles proposed by the LLM as a strong indicator of the model's performance. A thorough validation of Claude's annotations would require analysing all IS profiles and identifying potential errors, which is planned for future work.

The accuracy rating provides a direct evaluation of the Classification Task but serves only as an implicit and secondary measure for the Generation Task. A more direct assessment would require manual validation of the new dataset by domain experts, which has only been started here and is planned for future work.

4. Analysis and Discussion

In this section we provide some quantitative and qualitative analysis, as well as plausible interpretations of the output of the analysis we conducted on the Image Schema Catalogue and its LLM enrichment, and we provide explanation of charts and matrices shown in the followings.

These visualizations contribute to a comprehensive understanding of the relationship between human-generated and LLM-generated annotations, highlighting the strengths and limitations of the annotation model.

Match Counts Analysis   The most immediate and primary analysis we conducted was a bulk comparison between the original annotations and those generated by Claude 3.5.

This aims to examine the congruence between the initial human-assigned labels and the primary predictions of the LLM annotation system, assessing the model's capacity to accurately replicate original annotations. Furthermore, this allows us to quantify the occurrence rate of original annotations within the ranked LLM predictions, as shown in Table 1, and subsequently visualizing this distribution via a bar chart representation, shown in Figure 1.

Total Number of Entries	2559
Correct Annotations	2076
Correct Annotations in Pos. 1	1258
Correct Annotations in Pos. 2	489
Correct Annotations in Pos. 3	259
Correct Annotations in Pos. 4	64
Correct Annotations in Pos. 5	5
No Match	483

Table 1 Number of correct annotations and their position in Claude's relevance order.

The main goal of this work, as stated in the Introduction, is to enrich the IS Catalogue with more than one annotation per sentence. For this reason, in the analysis, we consider an overall "Correct Annotation" when the original annotation is present in the Image Schema Profile (the set of annotations)

provided by Claude. Overall, Claude correctly classifies 2075 sentences out of 2559, achieving 81% of accuracy. In more than half of cases (1258 sentences), the original classification is also Claude's top suggestion in the order of preference. Table 1 summarises the number of correct annotations and their position in the relevance orders provided by Claude.
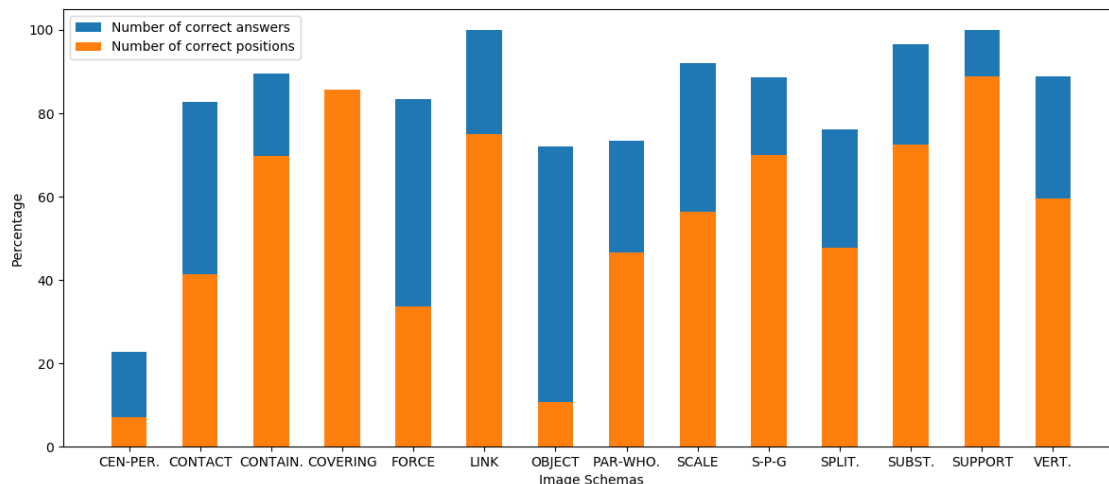

Figure 1: Match Counts Plot for each Image Schema included in the catalogue.

We also analysed the distribution of correct annotations across the different image schemas collected into the catalogue. The results are summarised in Figure 1: the highest accuracy rates in terms of correct annotation is obtained with the image schemas LINK and SUPPORT, both reaching 100% of accuracy. In contrast, the lower performance is reached with the image schema CENTER_PERIPHERY, being correctly identified only in the 22% of cases. Considering Claude's preference orders, the best performances are reached with SUPPORT and LINK, whereas the lower performances are related to the CENTER_PERIPHERY and OBJECT, the latter in particular being correctly identified in around 75% of cases, but only in 10% of cases in the correct position.

Comparative IS Distribution Analysis The methodology involves calculating the frequency distribution of each annotation type in the dataset. This comparative analysis offers insights into potential shifts in frequency and prioritization of specific annotations between human-generated and LLM-predicted labels. The bar charts in Figure 2 are employed to juxtapose the distribution of annotations between the original human-assigned labels (top bar chart) and the LLM annotations (center and bottom bar charts), identifying trends and discrepancies in annotation priorities between human-labeled and machine-labeled data.

In the case of the LLM annotations, we analysed here both the IS distribution considering the first element annotated by Claude (center) and the distribution of IS considering all annotation by Claude (bottom). As shown in Figure 2, the most used image schemas (both in the first position and generally, although with inverted order) are CONTAINMENT and SOURCE_PATH_GOAL. This finding is consistent with the original human annotation of the catalogue. Looking at the distribution of the human annotations, aside from CONTAINMENT and SOURCE_PATH_GOAL, the most frequent annotations are, in order, OBJECT, VERTICALITY, FORCE, and CENTER_PERIPHERY.

It is important to clarify that these 6 IS account for 91% of the entire catalogue. Notably, FORCE and VERTICALITY are also among the most common first annotations made by Claude, while OBJECT and CENTER_PERIPHERY are much less frequently used in the first position, ranking 7th and last for distribution, respectively. Considering the totality of Claude's annotation instead (Figure2, bottom chart), the distribution of OBJECT gains the fourth position, showing it is frequently used by Claude,

but not as its first preference.

The case of Center_Periphery is different, as its distribution is the lowest (excluding Claude's hallucination) even in the case of all Claude's annotations. These findings align with the accuracy analysis, which highlights Claude's difficulty in recognising the Center_Periphery image schema. More insights and data interpretations are provided ins Section 4.1.
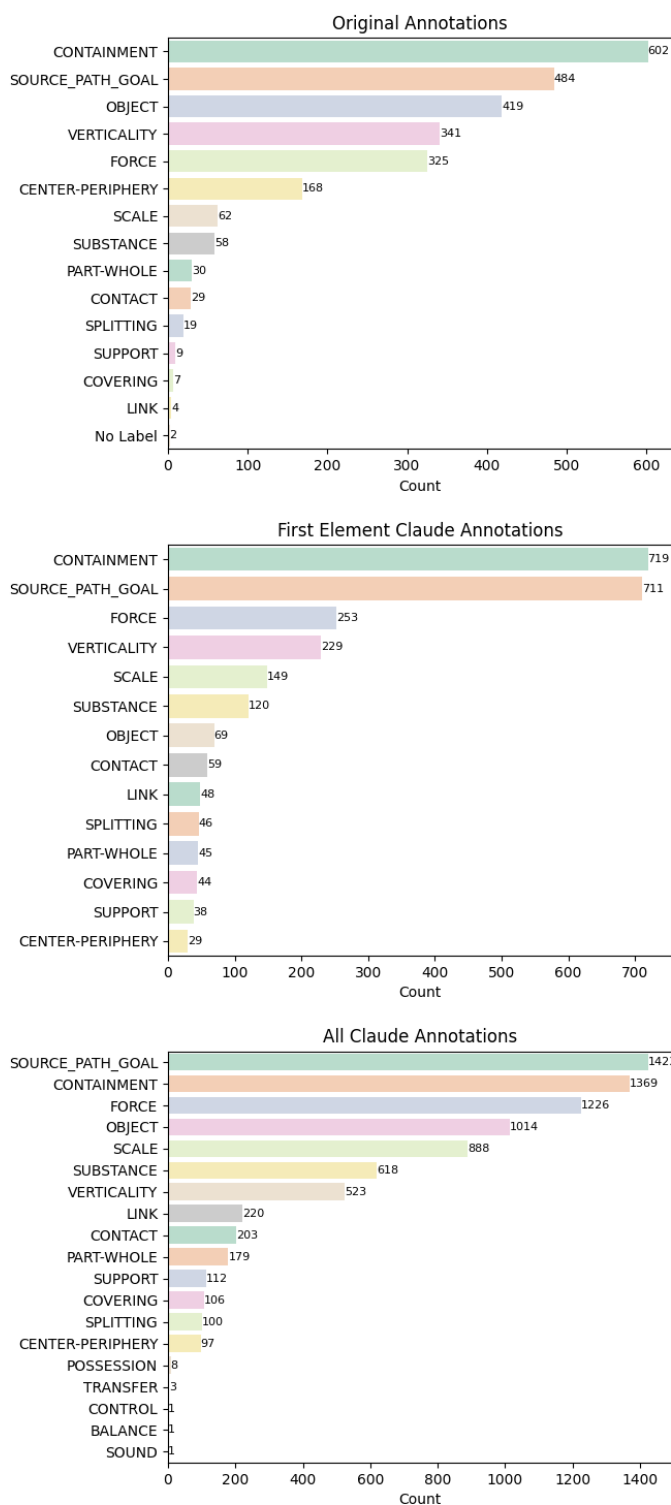
Co-Occurrence Matrix Analyses
The co-occurrence matrices are implemented to elucidate the interrelationships between annotation labels that frequently manifest in tandem. These matrices facilitates the identification of patterns and correlations within the LLM annotation sets that might otherwise remain obscured in discrete analyses. Each matrix cell quantifies the co-occurrence frequency of annotation pairs within a given set, revealing potentially significant associations. The procedural approach involves iterating through all LLM annotation lists, enumerating co-occurrences between annotation pairs, and presenting the results in a heatmap format.

We conducted two kinds of co-occurrence analysis. The first (cf. Figure 3) analyses the co-occurrences of Image Schemas for all Claude's annotations; the second (cf. Figure 4) repeats the analysis considering those cases for which the original annotation is matching the first element of the LLM IS profile.

As shown in Figure 3, the most frequent co-occurrences (over 500 instances) are Source_Path_Goal and Force, followed by Source_Path_Goal and Containment, Force and Containment, Source_Path_Goal and Object and Containment and Object. The matrix also

Figure 2: *Top*: Original Annotations; *Center*: First Element Annotations; *Bottom*: All Annotations

accounts for Claude's hallucinations. Figure 4 shows that the results for correct annotations align with those observed across all annotations.
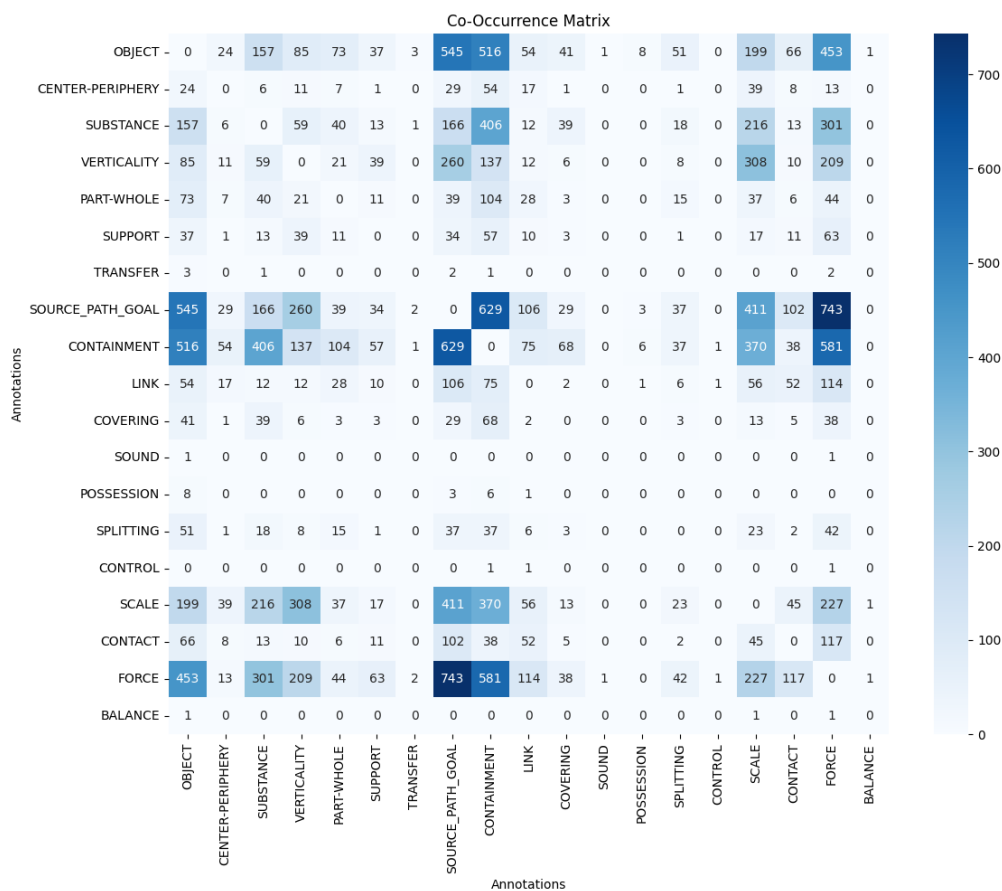
Co-Occurrence Matrix

Annotations	OBJECT	CENTER-PERIPHERY	SUBSTANCE	VERTICALITY	PART-WHOLE	SUPPORT	TRANSFER	SOURCE_PATH_GOAL	CONTAINMENT	LINK	COVERING	SOUND	POSSESSION	SPLITTING	CONTROL	SCALE	CONTACT	FORCE	BALANCE
OBJECT	0	24	157	85	73	37	3	545	516	54	41	1	8	51	0	199	66	453	1
CENTER-PERIPHERY	24	0	6	11	7	1	0	29	54	17	1	0	0	1	0	39	8	13	0
SUBSTANCE	157	6	0	59	40	13	1	166	406	12	39	0	0	18	0	216	13	301	0
VERTICALITY	85	11	59	0	21	39	0	260	137	12	6	0	0	8	0	308	10	209	0
PART-WHOLE	73	7	40	21	0	11	0	39	104	28	3	0	0	15	0	37	6	44	0
SUPPORT	37	1	13	39	11	0	0	34	57	10	3	0	0	1	0	17	11	63	0
TRANSFER	3	0	1	0	0	0	0	2	1	0	0	0	0	0	0	0	0	2	0
SOURCE_PATH_GOAL	545	29	166	260	39	34	2	0	629	106	29	0	3	37	0	411	102	743	0
CONTAINMENT	516	54	406	137	104	57	1	629	0	75	68	0	6	37	1	370	38	581	0
LINK	54	17	12	12	28	10	0	106	75	0	2	0	1	6	1	56	52	114	0
COVERING	41	1	39	6	3	3	0	29	68	2	0	0	0	3	0	13	5	38	0
SOUND	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
POSSESSION	8	0	0	0	0	0	0	3	6	1	0	0	0	0	0	0	0	0	0
SPLITTING	51	1	18	8	15	1	0	37	37	6	3	0	0	0	0	23	2	42	0
CONTROL	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0
SCALE	199	39	216	308	37	17	0	411	370	56	13	0	0	23	0	0	45	227	1
CONTACT	66	8	13	10	6	11	0	102	38	52	5	0	0	2	0	45	0	117	0
FORCE	453	13	301	209	44	63	2	743	581	114	38	1	0	42	1	227	117	0	1
BALANCE	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0

Annotations

Figure 3: Co-occurrence matrix on all Claude's annotations.

Confusion Matrix for Exact Match Accuracy   The confusion matrix serves as a critical evaluation tool for assessing the exact match accuracy between the original annotations and the first elements of the LLM annotations. This matrix quantifies both misclassifications and correctly predicted annotations. The rows represent the ground truth labels (original annotations), while the columns denote the predicted labels (primary elements of LLM annotations). By constructing and analyzing this confusion matrix, we evaluate the system's discriminative capabilities across annotation categories.

The Confusion matrix is shown in Figure 5: the most common misclassifications (>65 times) are between OBJECT and CONTAINMENT (113 times), SOURCE_PATH_GOAL and FORCE (110 times), CONTAINMENT and SOURCE_PATH_GOAL (81 times) and SOURCE_PATH_GOAL and CENTER_PERIPHERY (69 times).

4.1. Data Interpretation

Image Schemas classification   Overall, Claude performed relatively well on the Image Schema Catalogue, achieving an accuracy rate of 81%. Although Claude frequently identified the correct image schema among those composing the sentence IS profile, only in less than 50% of cases this was selected as the top most relevant choice.

The analysis we conducted in terms of accuracy shows the LLM's ability to identify the correct image schema with respect to human annotation. At the same time, by asking Claude to extend the annotation to more than one Image Schemas, its chances of guessing ('shooting in the heap') the correct image
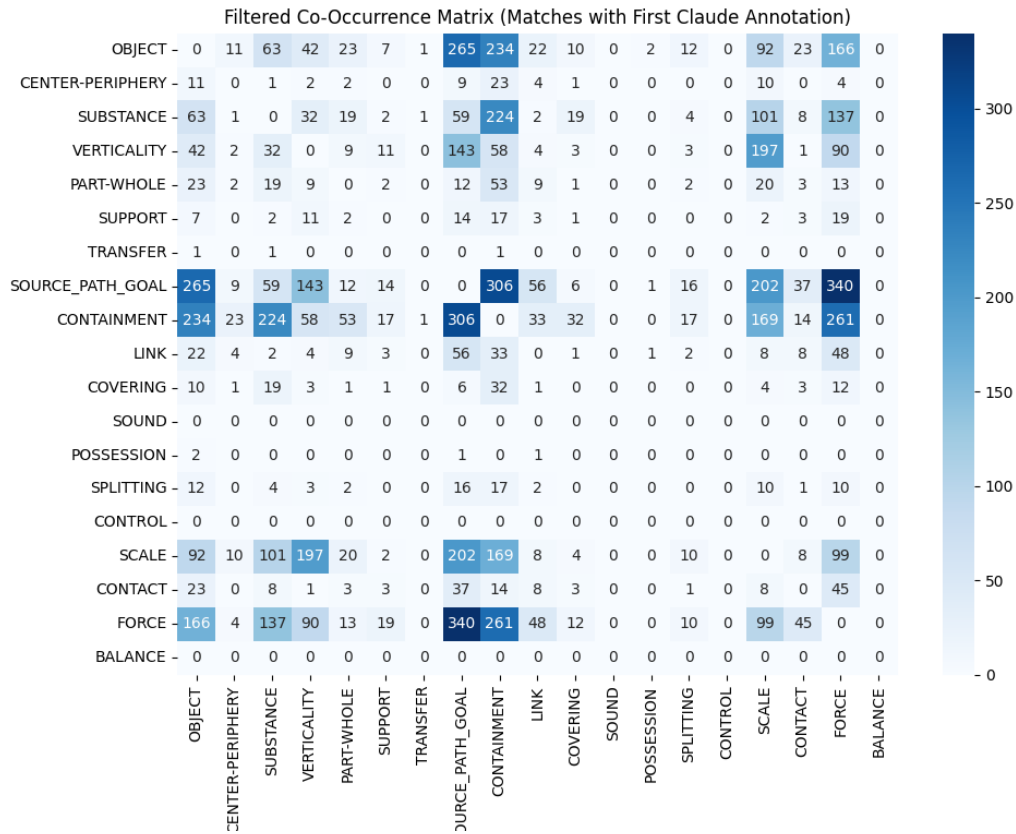
Figure 4: Co-occurrence matrix on first Claude's annotation.
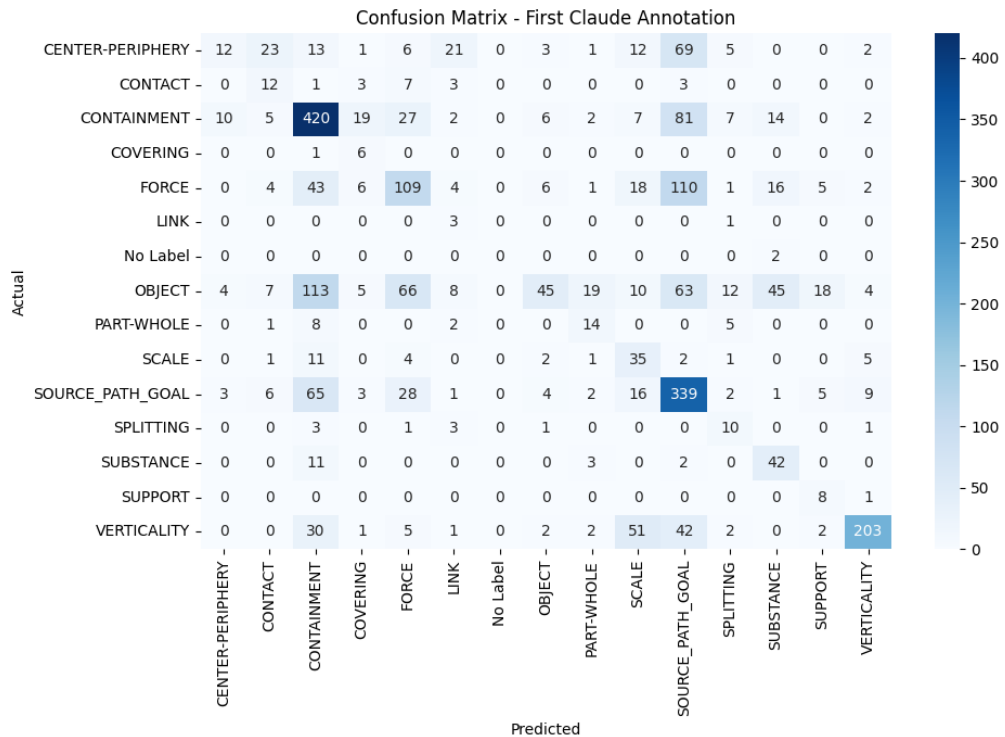

Figure 5: Confusion Matrix for Exact Match Accuracy

schema increase, given the limited number of image schemas to choose from. A complete and qualitative

validation of Claude's annotations for each entry would require (i) analysing all the IS profiles and (ii) checking for possible misinterpretations, and this is a matter of future work. Some examples, collected in Table 2, may however give some insight on the relevance and necessity of this work as well as the quality of the model in identifying meaningful Image Schemas.

ISC Sentence	Original Annotation	Claude's Annotations
Breaking social ties	Link	Splitting, Link, Force
Put more force into your punches.	Force	Force, Containment, Object
There's no way out, I have to do it.	Containment	Containment, Source_Path_Goal, Force

Table 2 Examples of Claude's annotations

The distribution of image schemas in the catalogue is uneven. For example, there are only 4 sentences for Link but 602 for Containment. Some results from the analysis (e.g. cf. Figure 1) should therefore be adjusted based on the frequency of each schema in the catalogue. For instance, Link and Support were correctly identified 100% of the time, but they appeared only 4 and 9 times, respectively, which distorts the sample to some extent. A similar case applies to Covering, which has only 6 entries in the catalogue.

More interesting cases are Object and Center_Periphery. As mentioned earlier, Object was correctly identified in around 75% of cases, though only in 10% of cases was it in the correct position (i.e., as the first choice). According to the frequency analysis, Object is also one of the most frequently used image schemas, annotated by Claude 1,014 times, but selected as the first preference in only 6% of cases. This pattern may suggest some uncertainty on Claude's part in applying the schema, although it could also reflect its broad role in conceptualising entities (see also below).

Claude struggled the most with the Center_Periphery schema, correctly identifying it in only about 20% of cases. Considering the frequency analysis, it emerges that it is also the least-used Image Schema, showing Claude's difficulty in recognising its pattern. Despite this challenge in the classification task, the performance in the generation task remained quite strong. A few examples follow.

Metaphorical Example: "She put the idea to the back of her mind." Literal Sentence: "She placed the book at the back of the shelf.".

Metaphorical Example: "These colors aren't quite the same, but they're close." Literal Sentence: "These buildings aren't quite adjacent, but they're near each other."

Metaphorical Example: "Stands to reason nobody'd dare to bomb us, because, we'd do the same to them." Literal Sentence: "The ball bounces back when it hits the wall."

Co-occurrences and confusion matrices The relevance of the valuable insights provided especially by the co-occurrence matrices and confusion matrix resides in the assumption that LLMs have ingested an enormous amount of data, and are for this reason the biggest commonsense approximate knowledge repositories we ever had. Some approaches [29] suggest that, although large language models are not embodied *per se*, they have processed such an amount of textual material to manifest, in their inductive reasoning - derived from statistical generalisation - a certain sparkle of embodiment.

From Claude's IS profile extraction several notable patterns emerge: Force and Source_Path_Goal show the highest co-occurrence rates, suggesting a strong conceptual link between force dynamics and directed motion or processes. Containment also frequently co-occurs with these schemas, indicating that bounded spaces often interact with forces and paths. Object schema has significant overlap with many others, which is unsurprising, given its fundamental nature in conceptualizing entities and the observations above. In fact, many spatial primitives co-participating to the realization of image schemas, can be conceived as instantiations of Object. Scale and Verticality show moderate co-occurrence, reflecting how vertical orientation often relates to scalar concepts. Interestingly, hallucinated images schemas like *Sound and *Possession have very low co-occurrence rates, suggesting they may represent

more specialized or isolated conceptual domains. The Substance schema shows notable co-occurrence with Containment, hinting at the frequent conceptualization of substances within containers.

It is also intriguing to note cases where expected co-occurrences are surprisingly low, despite the intuitive connections between certain image schemas. A notable example is the relationship between Support and Contact, which are instead prominent in knowledge representation's formalization of image schematic approaches to cognitive robotic applications [30, 31].

In many concrete scenarios, usually adopting a naive physics and reduced complexity spatio-relational dimensions representations—such as "an apple placed on a table"—these two schemas would typically appear together: the apple is in Contact with the table, and the table Supports it. However, when examining the co-occurrence matrix, the association between these two schemas is much lower than anticipated (they only co-occur 11 times), perhaps also due to the metaphorical nature of many of the phrases in the catalogue. Analogous is the case of Link and Source_Path_Goal as well as Link and Contact. Arguably, when a Link is present, it can be often envisioned a Path between two objects or situations (or vice versa) or at least some sort of connection that makes in Contact the two linked entities. However, within the catalogue, this correlation between image schemas occurs relatively a few times: only around a hundred times for Link and Source_Path_Goal and 52 for Link and Contact.

Co-occurrence and confusion measures interact with each other. The image schemas that are most frequently confused, according to the Confusion Matrix (Figure 5), also tend to co-occur frequently and are among the most commonly used across all annotations. This is true for pairs like Object and Containment, Source_Path_Goal and Force, and Containment and Source_Path_Goal, which are often confused but also co-occur frequently. In many of these cases, the correct schema is likely included in the annotations, just not in the first position.

A different scenario arises with Center_Periphery, which is often confused with Source_Path_Goal but co-occurs with it only in about one-third of cases. This suggests a genuine confusion between the two schemas. As mentioned above, Center_Periphery proved to be the most challenging schema for Claude. In addition to being confused with Source_Path_Goal, it was also frequently confused with Link (with which it co-occurs in nearly 50% of cases) and, more unexpectedly, with Contact (despite never co-occurring with it).

5. Conclusions and Future Work

In this paper, we have presented a novel approach to extending the Image Schema (IS) catalogue using Claude 3.5 Sonnet large language model. Our methodology has demonstrated remarkable efficacy, with the model successfully retrieving the original annotation in 81% of the annotated sentences, considering the whole set of Image Schemas extracted in the profile by the LLM. This very promising retrieval rate underscores the potential of leveraging advanced AI models in this kind of specific linguistic annotation tasks.

The enrichment of the IS catalogue resulting from this approach has been partially evaluated by domain experts, who have found it to be both sound and plausible. This preliminary validation lends credibility to our method and suggests that LLM-assisted extension of linguistic resources can produce high-quality synthetic data that align with expert knowledge, although further assessment is still needed to fully confirm this insight.

Given the current limitations in expanding the Image Schema Catalogue manually, our approach using large language models appears to be the most promising avenue for overcoming these constraints. The ability to rapidly process and annotate large volumes of text while maintaining accuracy and consistency offers a significant advantage over traditional methods.

However, this work also opens up several paths for future research. The first and most needed step would be a comprehensive expert evaluation: a full evaluation of the Extended IS Catalogue by domain experts is necessary to further validate and refine our approach. This will ensure the robustness and reliability of the expanded catalogue across diverse linguistic contexts. This step is needed both in relation to the IS Profiles annotated by Claude and for the newly generated sentences. Secondly,

topicalisation of IS: future works should focus on developing methods for annotating the specific textual chunks that evoke particular Image Schemas within sentences. This finer-grained analysis will provide deeper insights into how IS are linguistically realised. Finally, we envision a multimodal extension: with the advent of powerful multimodal models, there is potential to extend our approach beyond text. Incorporating visual and possibly auditory information could lead to a more comprehensive understanding of Image Schemas across different modalities of human conceptualization and cognition.

Acknowledgment

References

[1] J. M. Mandler, C. Pagán Cánovas, On defining image schemas, Language and Cognition (2014) 1–23. doi:`10.1017/langcog.2014.14`.

[2] L. Talmy, The fundamental system of spatial schemas in language, in: B. Hampe, J. E. Grady (Eds.), From perception to meaning: Image schemas in cognitive linguistics, volume 29 of *Cognitive Linguistics Research*, Walter de Gruyter, 2005, pp. 199–234.

[3] M. Johnson, The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason, The University of Chicago Press, Chicago and London, 1987.

[4] B. Hampe, Image schemas in cognitive linguistics: Introduction, From perception to meaning: Image schemas in cognitive linguistics 29 (2005) 1–14.

[5] M. M. Hedblom, O. Kutz, F. Neuhaus, Choosing the right path: image schema theory as a foundation for concept invention, Journal of Artificial General Intelligence 6 (2015) 21–54.

[6] J. Hurtienne, J. H. Israel, Image schemas and their metaphorical extensions: intuitive patterns for tangible interaction, in: Proceedings of the 1st international conference on Tangible and embedded interaction, 2007, pp. 127–134.

[7] J. Hurtienne, S. Huber, C. Baur, Supporting user interface design with image schemas: The iscat database as a research tool., in: ISD, 2022.

[8] G. Lakoff, M. Johnson, Metaphors we live by, University of Chicago press, 1980.

[9] G. Lakoff, M. Johnson, et al., Philosophy in the flesh: The embodied mind and its challenge to western thought, volume 640, Basic books New York, 1999.

[10] R. W. Langacker, Foundations of cognitive grammar: Theoretical prerequisites, volume 1, Stanford university press, 1987.

[11] R. W. Langacker, Cognitive grammar, Basic Readings 29 (2008).

[12] E. Dodge, G. Lakoff, Image schemas: From linguistic analysis to neural grounding, From perception to meaning: Image schemas in cognitive linguistics (2005) 57–91.

[13] B. Bennett, C. Cialone, Corpus guided sense cluster analysis: a methodology for ontology development (with examples from the spatial domain)., in: FOIS, 2014, pp. 213–226.

[14] A. Cienki, Image schemas and gesture, From perception to meaning: Image schemas in cognitive linguistics 29 (2005) 421–442.

[15] M. M. Hedblom, O. Kutz, T. Mossakowski, F. Neuhaus, Between contact and support: Introducing a logic for image schemas and directed movement, in: Conference of the Italian Association for Artificial Intelligence, Springer, 2017, pp. 256–268.

[16] G. Righetti, D. Porello, N. Troquard, O. Kutz, M. M. Hedblom, P. Galliani, Asymmetric hybrids: Dialogues for computational concept combination, in: Formal Ontology in Information Systems, IOS Press, 2021, pp. 81–96.

[17] G. Righetti, O. Kutz, The moving apple: An image-schematic investigation into the leuven concept database, in: Proceedings of The Seventh Image Schema Day co-located with The 20th International

Conference on Principles of Knowledge Representation and Reasoning (KR 2023), Rhodes, Greece, September 2nd, 2023, CEUR-WS, 2023.

[18] S. De Giorgis, A. Gangemi, D. Gromann, Imageschemanet: Formalizing embodied commonsense knowledge providing an image-schematic layer to framester, Semantic Web Journal forthcoming (2022).

[19] M. Hedblom, O. Kutz, R. Penaloza, G. Guizzardi, et al., What's cracking? how image schema combinations can model conceptualisations of events, in: CEUR WORKSHOP PROCEEDINGS, volume 2347, CEUR-WS, 2019.

[20] M. M. Hedblom, F. Neuhaus, T. Mossakowski, The diagrammatic image schema language (disl), Spatial Cognition & Computation (2024) 1–38.

[21] A. Papafragou, C. Massey, L. Gleitman, When English proposes what Greek presupposes: The cross-linguistic encoding of motion events, Cognition 98 (2006) B75–B87.

[22] J. A. Prieto Velasco, M. Tercedor Sánchez, The embodied nature of medical concepts: image schemas and language for pain., Cognitive processing (2014). doi:10.1007/s10339-013-0594-9.

[23] D. Gromann, M. M. Hedblom, Body-mind-language: Multilingual knowledge extraction based on embodied cognition, in: AIC, 2017, pp. 20–33.

[24] D. Gromann, M. M. Hedblom, Kinesthetic mind reader: A method to identify image schemas in natural language, in: Proceedings of Advancements in Cogntivie Systems, 2017.

[25] L. Wachowiak, D. 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, 2022, pp. 5571–5581.

[26] S. De Giorgis, A. Gangemi, Introducing odin: Ontological design grounded in image-schematic knowledge., in: WOP@ ISWC, 2022.

[27] M. M. Hedblom, O. Kutz, R. Peñaloza, G. Guizzardi, Image schema combinations and complex events, KI-Künstliche Intelligenz 33 (2019) 279–291.

[28] T. Oakley, Image schemas, The Oxford handbook of cognitive linguistics (2007) 214–235.

[29] S. Nolfi, On the unexpected abilities of large language models, Adaptive Behavior (2023) 10597123241256754.

[30] M. Pomarlan, S. De Giorgis, M. M. Hedblom, M. Diab, N. Tsiogkas, Thinking in front of the box: Towards intelligent robotic action selection for navigation in complex environments using image-schematic reasoning., in: JOWO, 2022.

[31] M. Pomarlan, S. De Giorgis, R. Ringe, M. M. Hedblom, N. Tsiogkas, Hanging around: Cognitive inspired reasoning for reactive robotics, in: Formal Ontologies for Information Systems (FOIS) 2024, 2024.