Smith ScholarWorks

Computer Science: Faculty Publications

Computer Science

2022

Novel Primitive Decompositions for Real-World Physical Reasoning

Mackie Zhoou
Smith College

Bridget Duah
Smith College

Jamie C. Macbeth
Smith College, jmacbeth@smith.edu

Follow this and additional works at: https://scholarworks.smith.edu/csc_facpubs

Part of the Computer Sciences Commons

Recommended Citation

Zhoou, Mackie; Duah, Bridget; and Macbeth, Jamie C., "Novel Primitive Decompositions for Real-World Physical Reasoning" (2022). Computer Science: Faculty Publications, Smith College, Northampton, MA.
https://scholarworks.smith.edu/csc_facpubs/371

This Article has been accepted for inclusion in Computer Science: Faculty Publications by an authorized administrator of Smith ScholarWorks. For more information, please contact scholarworks@smith.edu

Novel Primitive Decompositions for Real-World Physical Reasoning

Mackie Zhou MZHOU@SMITH.EDU
Bridget Duah BDUAH@SMITH.EDU
Jamie C. Macbeth JMACBETH@SMITH.EDU
Smith College, Northampton, MA, USA

Editor: Kristinn R. Thórisson

Abstract

In this work, we are concerned with developing cognitive representations that may enhance the ability for self-supervised learning systems to learn language as part of their world explorations. We apply insights from in-depth language understanding systems to the problem, specifically representations which decompose language inputs into language-free structures that are complex combinations of primitives representing cognitive abstractions such as object permanence, movement, and spatial relationships. These decompositions, performed by a system traditionally called a *conceptual analyzer*, link words with complex non-linguistic structures that engender the rich relations between language expressions and world exploration that are a familiar aspect of intelligence.

We focus on improving and extending both the Conceptual Dependency (CD) representation system, its primitive decompositions, and its conceptual analyzer, choosing as our corpus the ProPara ("Process Paragraphs") dataset, which consists of paragraphs describing biological, chemical, and physical processes of the kind that appear in grade-school science textbooks (e.g., photosynthesis, erosion). In doing so, we avoid the significant challenges of decomposing concepts involving communication, thought, and complex social interactions. To meet the challenges of this dataset, we contribute a *mental motion pictures* representation system with important innovations, such as using image schemas in place of CD primitives and decoupling containment relationships into separate primitives.

Keywords: Self-Supervised Learning; Natural Language Understanding; Conceptual Primitives; Conceptual Representations

1. Introduction

Although deep learning systems may be evolving cognitive representations from data, work on both adversarial examples (Zhang et al., 2020) and work to evolve new benchmarks for language models (Srivastava et al., 2022) indicates they may not be the kinds of representations that humans use when performing the same tasks. As researchers are now scaling large language models to more than one trillion parameters to achieve superhuman performance (Rajbhandari et al., 2020; Fedus et al., 2022), the models may only be exploiting patterns embedded in the datasets, perhaps in the behavior patterns of crowdworkers providing dataset items, or in the narrow definition of the task itself. Therefore, scientific exploration of cognitive representation systems and imagery representation systems remains relevant.

This paper continues work on cognitive representation systems which represent concepts as decomposed structures that are complex combinations of abstract primitives. A major

© 2022 M. Zhou, B. Duah & J.C. Macbeth.

challenge of this work is to, as much as possible, make these representations language free, so that common-sense reasoning can proceed through flexible, complex, deeply-nested symbolic structures of mental imagery, separate from the variety of ways they can be expressed linguistically. Common-sense reasoners benefit from flexible representation systems because they allow for rich connections between the current image in memory and knowledge structures, whether the goal is understanding, problem solving, or another task. At the same time, fostering a division between linguistic and non-linguistic representations requires the development of complex systems to "translate" between the two. Advances in flexible, human-like knowledge representations and meaning representations will enhance self-supervised learning by enabling a rich array of mappings between new experiences and structures in short-term and long-term memory. Self-supervised learning can be further enhanced when language interaction itself is embedded in the environment through communication with a parent or teacher agent.

Based on the English Language Interpreter (ELI, Birnbaum and Selfridge, 1981), we present a prototype of a new conceptual analyzer that decomposes a simple natural language expression involving physical objects into a *mental motion picture* representation which breaks down a short episode of a story into consecutive critical moments, imitating continuous frames of a film or video clip. In each frame, *mental maps* capture relationships between the physical objects involved in the stories from different dimensions. By recording mental maps at each critical moment, the system keeps track of the evolution of physical objects mentioned in an expression, facilitating the translation of a text to a non-linguistic representation.

This paper is organized as follows: The first section briefly discusses a dataset (ProPara) that we use as a corpus of language for developing and testing our representational systems. Next we describe the mental motion pictures representational system and systems for transforming language expressions into the representation. After a detailed example of the systems at work and a discussion of them, we touch on related work and conclude with ideas for further exploration.

2. ProPara

The ProPara ("Process Paragraphs", Dalvi et al., 2018) dataset consists of 488 paragraphs (3,303 sentences), describing biological, chemical, and physical processes of the kind that appear in grade-school science textbooks (e.g., photosynthesis, erosion). The paragraphs typically have between five and eight simple sentences and annotations designed to test comprehension and question answering about how entities move and change during a process.

The following is an example paragraph from ProPara that is a response to the prompt "How do caverns form?"

> Limestone lies under the soil. Rain picks up carbon dioxide as it falls to earth. The rain falls on the soil over the limestone. The carbon dioxide in the rain washes through the soil. The carbon dioxide turns into acid. The acid in the rain gets to the limestone below the soil. The acid dissolves the limestone. Acid continues to erode the limestone with more rain over time. The eroded limestone sometimes forms caves.

The ProPara dataset is ideal for evaluating our conceptual analyzer and mental motion picture representation system because the processes described generally involve only physical acts and events, allowing for focus on physical primitives and their corresponding language without requiring consideration for decomposing concepts involving communication, thought, and complex social interactions (Macbeth, 2020).

3. Mental Motion Pictures

The meaning representation system that we have devised to represent understanding of ProPara sentences and paragraphs is a mental imagery system with world objects, simple spatial maps (graphs) of spatial relationships between objects, and event primitives which represent changes in the states of objects, or changes in their spatial relationships.

3.1. Structure of Mental Motion Picture Representations

We structure a mental motion picture (MMP) to represent the passage of time as a sequence of frame objects. The mental motion picture has a pointer to the most recent frame. In order to record as much information as possible, a composer of mental motion pictures writes incidents and primitive acts to the most recent frame, and then the pointer moves to the next frame as the story progresses. We can think of a mental motion picture as a timeline, where we record activities that happened in a moment and move to the next moment when we are done collecting information about the current moment. Each frame consists of three mental maps that help describe incidents that happened in that specific frame: a containment map, a position map, and a contact map. We will elaborate on mental maps and the frame functionality in the following sections.

3.2. Mental Maps

In a frame, each mental map is a graph object where vertices are objects that appear in the story, and edges between vertices indicate connections between objects. The mental maps depict three different kinds of connections or relationships between physical objects using the containment map, the position map, and the contact map.

3.3. The Containment Map

The containment map aims to depict containment relationships between objects. When object B goes inside of object A, we say that A contains B. A real-world example would be this: a pen is put into a pencil case, so we say that the pencil case now contains the pen.

A containment relationship is directed and must not go both ways. In other words, if A contains B, then we know for sure that B does not contain A. In this case, an edge [A, B] means object A contains object B.

3.4. The Position Map

The goal of the position map is to record the position of an object with respect to the center of the earth. In other words, the closer the object is to the center of the earth, the lower the altitude of an object. Again, the spatial relationship between objects is directed. For the two physical objects A and B, we decide that A can either be above B or below, but not both. An edge in the position map must be directed as well. We define that an edge [A, B] in a position map means object A is closer to the center of the planet than object B is (i.e. M is lower than N).

We chose the "center of earth" perspective because we thought it would make it easier to construct an image structure, especially in cases when the paragraphs describe processes that involve objects and acts that occur both above and below the surface of the earth. Using an more egocentric view, for example, recording the position of an object with respect to the surface of the earth, requires "falling" to be a movement away from or toward the surface of the earth depending on the object's location. This, in our view, would add complexity to composing the representations. One drawback of our system, however, is that it requires the creation of the center of the earth as an object in the representations even when it is not explicitly mentioned.

3.5. The Contact Map

The contact map records if two objects involved in the story are touching one another. A touching relationship is undirected, meaning that if A touched B, B must also touch A. Therefore, edges in the contact map are undirected. An edge [A, B] means that these two objects A and B are touching each other.

All three maps must exist in any frame. If a new physical object appears in the story, the object will be added to all three maps. However, in terms of edges, the three maps are independent of each other. For example, if a containment relationship changes between two objects, an edge will be removed or added to the containment map, but both the position map and the contact map should be unaffected. There may be cases where a new edge is added to more than one map at once, but we still treat maps and edges independently.

3.6. Frames

A frame is a node of the mental motion picture (linked list). Besides that, a frame is capable of inheriting objects, edges, and primitive acts from the frame that precedes it. Unless a physical object transforms into another object, gets destroyed, or disappears from the world of the story in the timespan of the current frame, it will be copied from the current frame to the next frame. The same principle applies to primitive acts and edges in the three maps as well.

3.7. Primitives

The mental motion pictures representation is inspired by Conceptual Dependency (CD, Schank, 1972, 1975) which has the conceptual primitive PTRANS to represent the movement of an object, and by work that crosses CD primitives with image schemas (Macbeth et al., 2017; Gromann and Macbeth, 2018; Macbeth and Gromann, 2019). We use the CD PTRANS primitive, which represents movement of an object. In CD the PTRANS primitive has conceptual "cases", "to" and "from." We slightly changed the imagery "semantics" of the PTRANS primitive so that the "to" case of PTRANS means that the movement is in the direction of the object in the "to" case, but may not ultimately arrive there. For example, we represent "fall" as an object PTRANSing "to the center of the earth" to indicate that the object is heading "down" even if it is underground, even though its motion will probably be stopped before reaching the center of the earth. Similarly, the "from" case of a PTRANS indicates that the movement is away from the object specified, but may not have started there.

For this reason we claim that, for words and prepositional phrases involving PTRANS, our system may provide a more detailed representation than a traditional CD conceptual analysis. Depending on the input sentence, with the help of the frame functionality, the action can be broken down to multiple frames and thus offer a more detailed representation of object movement. For example, it is possible that an object is moving towards a direction but stops before hitting that place (a ball falls towards the center of the planet but stops when it hits the ground; the ball never reached the center of the planet). In this case our system outputs a more detailed analysis of the text than previous conceptual analyzers can. We also found the need to create a new primitive PSTOP which represents a mental image of an object stopping its motion and becoming still.

Instead of using the CD primitives INGEST and EXPEL, we have replaced them with further decompositions which combine movement (PTRANS) with a change in containment relationship, as suggested by prior comparisons and merges of CD with Lakoff-Johnson-Mandler image schemas (Macbeth et al., 2017). INGEST-like situations (one object going into another object) are represented as PTRANS combined with a new containment relationship being created, while EXPEL-like situations (one object coming out of another object) are represented as PTRANS combined with the end of a containment relationship.

4. The Conceptual Analyzer

Mental motion pictures constitute a meaning representation system for the kind of process-oriented natural language in our corpus, and an instance of a mental motion picture is meant to be the final product of a conceptual analysis, a process of "translation" from language into a meaning representation structure instance. The conceptual analyzer's processing is meant to simulate the human cognitive process of composing a complete imagery of the situation from partial imagery structures invoked by individual words. After reading or hearing a word or phrase, human understanders have a vague expectation of the kinds of words or phrases that should follow. For example, if we read the incomplete sentence "Mary uses," we expect that there will be a noun-phrase after the word "uses" and that noun-phrase describes the object that is used by a person named "Mary." Syntactic structures can also invoke mental motion picture structures: for example, the start of a new sentence in a paragraph is a typical trigger for creating a new frame. Our conceptual analyzer is an "expectation"-based conceptual analyzer modeled on an earlier conceptual analyzer called the English Language Interpreter (ELI, Birnbaum and Selfridge, 1981). We adopt the same terminology as is used in prior descriptions of a "micro" version of ELI (Birnbaum and Selfridge, 1981).

Expectation-based analyzers read the input word by word and create partial or incomplete non-linguistic conceptual structures corresponding to particular words, and instructions for completing the structure with conceptual structures instantiated and built through the expectation processes of other words in the utterance. In our system, as with ELI, expectations are represented by “requests.” The analyzer’s state consists of a stack that it uses to organize and process request structures and a set of “global” variables and registers that can be assigned by triggered requests. The registers comprise a working memory and are used in performing a variety of functions, such as keeping track of the current state of the syntax parse (e.g. PART-OF-SPEECH, and SUBJECT), or storing the current mental motion picture structure that is being constructed.

4.1. The Conceptual Lexicon

The conceptual analyzer also has a *conceptual lexicon*, which contains the imagery and expectation structures to instantiate for its word-by-word processing. The conceptual lexicon contains *packet* structures corresponding to words. A conceptual lexicon entry consists of a single packet, and is effectively a “small program” comprised of a list of *requests*. Each request in a packet consists of a TEST statement with conditional expressions that control when the request is triggered and executed. Requests also have ASSIGN and NEXT-PACKET statements that comprise the execution. An ASSIGN statement has a list of variables and the values they are to be assigned to when the request is triggered, and a NEXT-PACKET statement contains another packet which is to be pushed onto the stack and processed after the ASSIGNments are performed. ASSIGN statements may be used to instantiate new mental motion picture structures, or fill in “gaps” in previously instantiated structures. All parts of a request are optional. A missing TEST or a TEST with condition “true” means that the request always triggers.

For example, in the word packet for “FALLS,” there is one request. The request has a TEST with condition “true” so the request is always ready to be triggered and returned. The request assigns “PART-OF-SPEECH” to be “verb.” Assignments to PART-OF-SPEECH allow the analyzer to trigger later actions based on the expectation of a particular syntactic phrase structure item. The request also performs an assignment that adds a “PTRANS from an unknown place to the EARTH” to the current frame. Lastly, a NEXT-PACKET in the request pushes another packet onto the stack that specifies that it expects the word “ON” or “FROM” in the analysis process, filling the TO and/or FROM “gaps” of a PTRANS.

4.2. The Conceptual Analysis Process

When it process a word, the analyzer first finds its entry in the conceptual lexicon and, if it exists, it places the packet corresponding to that word on top of a stack of previously loaded packets (words without entries are ignored). The analyzer then examines the packet that has just been placed on top of the stack for any requests that could be triggered (for example by the value of an analyzer variable such as PART-OF-SPEECH matching the trigger conditions). If one of the requests in the top packet is triggered, it is popped off of the stack and executed.

The execution may, through a NEXT-PACKET in the request, cause another packet to be pushed on the stack just after the current packet has been removed. Regardless of whether another packet is pushed on, the packet that is newly the top of the stack is examined for a triggered request. The process of examining, triggering, and executing requests from the packet at the top of the stack continues until no request is triggered or the stack is empty. When the stack triggering and execution halts, the analyzer moves on to the next word in the input. This process is identical to that of Micro ELI (Birnbaum and Selfridge, 1981).

4.3. Conceptual Analyzer Augmentations

In most cases, a word packet stays on the stack until the expectation formed by reading this word is satisfied (in other words, the request embedded in that word packet is triggered). It is often the case that after reading a word, we expect the next word to have some specific features. As in the example mentioned above, after reading the word "uses," we expect to see an object (very likely a noun phrase) immediately after. In order to realize that, we have given each word packet a NEXT-PACKET attribute in which we store some more word packets. When an expectation is satisfied (i.e. a request is triggered), word packets stored in the NEXT-PACKET attribute of the current packet will be automatically pushed to the stack. The whole process is meant to resemble a human thought process—we read a word and subconsciously form expectations of what will come next, and if an expectation is satisfied, we form even more expectations and hope the next phrase in the text will satisfy those newly formed expectations.

However, in certain situations it is useful for a packet to remain on the stack after it is triggered (what we have dubbed "keepstack" functionality). This allows our conceptual analyzer to represent parallel expectations. For example, after hearing a particular verb, there may be expectations for many different kinds of prepositional phrases which could appear in any order. In this case, it allows packets having expectations for different prepositions to remain on the stack after any of those prepositions has been processed. A packet with its "keep" flag set to true is not removed when a request in it is triggered (but can be removed in other circumstances).

5. An Example

We now present a detailed description of the conceptual analyzer building a mental motion picture representation for the example sentence: "The rain falls on the soil from the cloud." This sentence is reminiscent of sentences found in ProPara.

1. An analyzer is initialized. A first empty frame is created. All three maps are empty. No primitive act yet.

2. Input the example sentence. The sentence is tokenized. By default, a START flag will be added to the sentence as the first word to direct the analyzer to start analyzing, so the resulting list is ["*START*", "THE", "RAIN", "FALLS", "ON", "THE", "SOIL", "FROM", "THE", "CLOUD"]. The START flag entry in the lexicon places a number of requests on the stack that represent the subject-verb-object syntax of English sentences. (By default, the analyzer assumes that the structure of a simple English sentence is SUBJECT-VERB-OBJECT.)

3. Read the word "*START*". The analyzer will start the analysis. *START* places packets on the stack that represent general expectations about the structure of an English sentence. For a simple English sentence involving physical objects, we assume

the sentence will follow the subject-verb-object structure. Therefore, in the "NEXT-PACKET" attribute of the word packet for "*START*" (call it "packet A"), we store a word packet that has only one request which will be triggered if encountering a noun-phrase later. Further, in the "NEXT-PACKET" attribute of packet A is another packet (packet B) that has only one request which will be triggered if encountering a verb in the later parsing process. We may edit/add word packets in the NEXT-PACKET attribute of "*START*" to deal with more complex sentence structures.

4. Read the word "THE." Currently the analyzer does not have any structures for dealing with articles.

5. Read the word "RAIN."

   The "CD" attribute of the analyzer is then set to "RAIN" and "PART-OF-SPEECH" to "noun-phrase". The object "RAIN" is added to all three maps of the first frame.

   The analyzer realizes that the current PART-OF-SPEECH is a noun-phrase and assumes that this is the subject of the sentence. Therefore, the SUBJECT attribute of the analyzer is set to the current CD, which is RAIN.

6. Read the word "FALLS."

   The "CD" attribute of the analyzer is then set to "FALLS" and "PART-OF-SPEECH" to "verb."

   According to the word entry for the word "FALLS" in the lexicon, this verb denotes that something is moving from somewhere to the direction of the center of the planet. Therefore, a PTRANS will be added to the first frame.

   FALLS also adds a "keepstack" packet to the stack representing the expectations for prepositional phrases that add detail to the mental motion picture. There may be multiple prepositional phrases in the sentence that are expected after FALLS.

   For the word "FALLS," after reading it, we naturally wonder where the object falls from and where it is going to stop at. The preposition "FROM" can be used to elaborate on the former, and the prepositions "TO" or "ON" can be used to specify the latter. For this reason, in the NEXT-PACKET attribute of the word packet for "FALLS," we can have a packet that contains multiple requests for different prepositional phrases (one request for "FROM", one for "ON", and maybe one for "TO"), and we want to set the "keep" flag of the next packet to be true. As a result, once the packet containing these requests is put on the stack, it won't be removed until all prepositional phrases we expect to see have shown up. These prepositional phrases can appear in any order and the result of analysis won't be affected.

7. Read the word "ON"

   The "CD" attribute of the analyzer is then set to "ON" and "PART-OF-SPEECH" to "preposition."

   A preposition "ON" after the word "FALLS" denotes that the subject stops moving as it touches something. The analyzer realizes that something new is happening and will advance the frame (create Frame 2 and copy everything in Frame 1 to Frame 2).

Now, in Frame 2, a PSTOP primitive act is added to show that the rain stops moving from here on.

8. Read the word "THE." Currently there are no conceptual lexicon entries for articles, so the analyzer moves on to the next word.

9. Read the word "SOIL."

   The "CD" attribute of the analyzer is then set to "SOIL" and "PART-OF-SPEECH" to "noun-phrase." The object "SOIL" is added to all three maps of the second frame.

   The lexicon tells the analyzer that if there is a noun after "FALLS ON," that noun is where the object stops moving. Therefore, the analyzer will update the PTRANS and specify that the "RAIN" is moving towards the "SOIL". The update will be made to both Frame 1 and Frame 2.

   The lexicon also tells the analyzer that, in terms of spatial relationship, the subject is above the noun after "ON." Therefore, an edge (SOIL, RAIN) will be added to the spatial map of the current/second frame. Because the RAIN stops as it is touching the SOIL, an edge (SOIL, RAIN) is also added to the contact map of the current/second frame.

10. Read the word "FROM"

    The "CD" attribute of the analyzer is then set to "FROM" and "PART-OF-SPEECH" to "preposition."

11. Read the word "THE."

12. Read the word "CLOUD."

    The "CD" attribute of the analyzer is then set to "CLOUD" and "PART-OF-SPEECH" to "noun-phrase." The object "CLOUD" is added to all three maps of the second frame.

    The lexicon tells the analyzer that if there is a noun after "FALLS FROM," that noun indicates the location where the object starts moving. Therefore, the analyzer will update the PTRANS and specify that the "RAIN" is moving from the "CLOUD". The update will be made to both Frame 1 and Frame 2. (Although it would be sensible to also add a (RAIN, CLOUD) edge to the position map indicating that the RAIN is closer to the center of the earth than the CLOUD as it is falling, the current lexicon entry for "FALLS FROM" does not allow for this inference to be made.)

A diagram of the state of the conceptual analyzer, with its mental motion picture structure, after analyzing the sentence, is shown in Figure 1.

6. Discussion

Presentations of Conceptual Dependency in the classical AI literature (Schank, 1975) have a strong emphasis on the primitive acts and events that largely represent meanings through state change. In this exploration of conceptual analysis of sentences in process paragraphs,

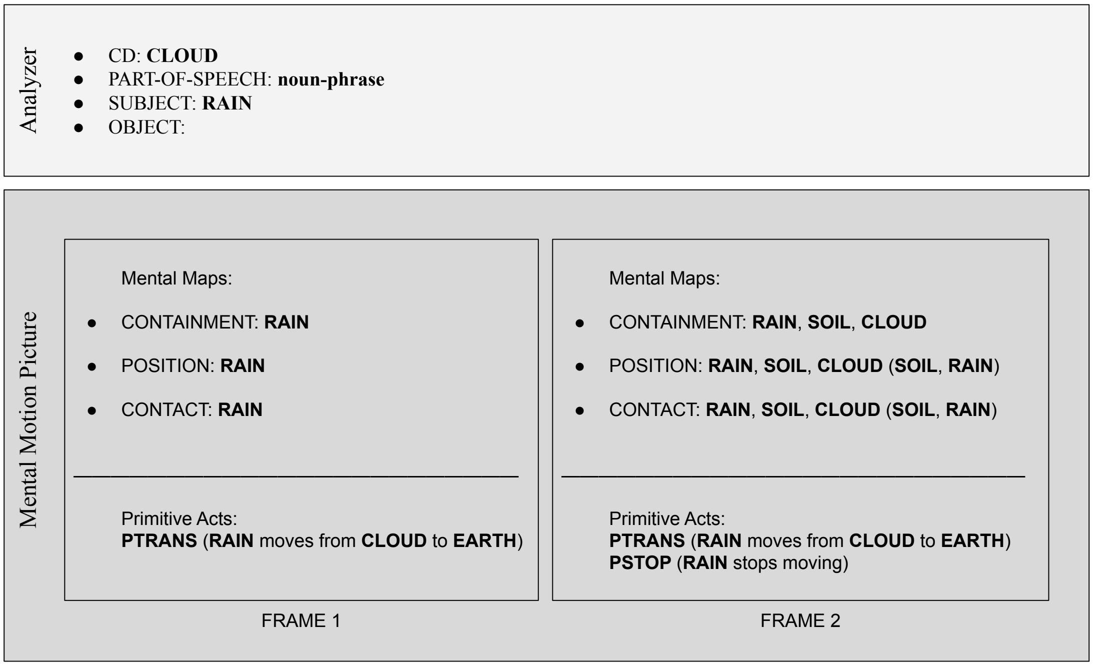
Analyzer
• CD: **CLOUD**
• PART-OF-SPEECH: **noun-phrase**
• SUBJECT: **RAIN**
• OBJECT:
Mental Motion Picture
Mental Maps:
• CONTAINMENT: **RAIN**
• POSITION: **RAIN**
• CONTACT: **RAIN**
Primitive Acts:
PTRANS (**RAIN** moves from **CLOUD** to **EARTH**)
FRAME 1
Mental Maps:
• CONTAINMENT: **RAIN**, **SOIL**, **CLOUD**
• POSITION: **RAIN**, **SOIL**, **CLOUD** (**SOIL**, **RAIN**)
• CONTACT: **RAIN**, **SOIL**, **CLOUD** (**SOIL**, **RAIN**)
Primitive Acts:
PTRANS (**RAIN** moves from **CLOUD** to **EARTH**)
PSTOP (**RAIN** stops moving)
FRAME 2

Figure 1: The state of the conceptual analyzer, with its mental motion picture structure, after analyzing the sentence "The rain falls on the soil from the cloud."

we found it necessary to place a greater emphasis on micro-primitives to decompose "static" unchanging situations, largely through existence of mentioned objects, spatial relationships, and containment relationships. For example, the "the rain falls on the soil" ends with a frame in which the rain is at rest. Although there are rare examples of CD structures in the literature which have a special notation to represent the end of a PTRANS act and thus the resulting static situation, we found these situations were so common that we introduced a PSTOP primitive. This embodies an emphasis in our work on representing static situations easily and elegantly.

Another aspect of representing static situations is in spatial relationships of existing objects. With the help of the containment relationship, we can infer relative sizes of objects that have containment relationships. We assume that the container is larger in size than the object that is inside of the container. With more multiple containment relationships that share the same "container" or "containee", we can even sort the objects by their sizes. For example, if a pen is put into a pencil case, we will create a new edge in the containment map (pencil case, pen), which indicates that "the pencil case contains the pen." This edge in the containment map infers that the size of the pen is smaller than that of the pencil case. If we have another edge (backpack, pencil case) in the containment map, we can say, in terms of sizes, backpack > pencil case > pen. The containment map provides this affordance of making sense of or sorting objects by their relative sizes, which is a unique feature of our system.

7. Related Work

Hunter et al. (2008) use a method similar to conceptual analysis called Direct Memory Access Parsing to analyze texts about biological processes, specifically protein transport, protein interactions, and gene expression. Dyer's Parser, which was the conceptual analyzer for Michael Dyer's BORIS in-depth story understander (Dyer, 1982) took an alternative approach from ELI in using "demons" instead of a stack-based approach.

Related work by Jackendoff (1983), Wierzbicka (1996), and Wilks and Fass (1992) explores primitive decomposition systems for both linguistics and natural language understanding. Important related work has explored the region connection calculus, a logical system for qualitative spatial representation and reasoning (Cohn et al., 1997). More recently, Thórisson et al. (2014) demonstrated a system which learns to communicate with little-to-no up-front knowledge by incrementally producing models of causal relationships in its observations, while Zeng and Davis (2021) explore an implementation of an open-world reasoner for a toy microworld of blocks and containers that can be loaded unloaded, sealed, unsealed, carried and dumped.

8. Conclusion

Our system attempts to compose an understanding of a text solely using the structures provided in the conceptual lexicon. The lexicon needs to specify as much information about the word it represents as possible so that the result of the analysis makes sense. However, in real life situations, humans perform commonsense inferences and introduce objects and facts that were not explicitly mentioned as part of understanding the meaning behind an utterance.

For example, when an English speaker reads or hears the verb "to fall" in a sentence or phrase it becomes natural to think that something is moving towards the earth or towards its center due to gravity, even though the earth is not mentioned. Our conceptual analyzer, through its conceptual lexicon, does create this context (via the "to" case of a PTRANS). In future works, we may devise a commonsense knowledgebase outside of the conceptual lexicon but with structures similar to it that can work alongside it by performing inferences which further flesh out the mental motion picture. Further research will explore where to draw the line between "knowledge" that ends up in the conceptual analyzer lexicon, and general knowledge in a commonsense knowledgebase.

One serious limitation of the simplification applied by the mental motion pictures system is in the way it represents objects. The presented mental moving pictures representation can decompose situations involving movement and containment, and in some cases, containment relationships can be used to infer shapes. For example, if a sheet of paper goes into a folder, we can infer that not only the paper and the folder shares a containment relationship (the folder contains the sheet of paper), but also the two objects share the same shape (both are thin and flat). However, if we think of the case of putting a cake into a refrigerator, this type of inference is not valid. Important future work will involve devising new primitives that allow the decompositions of shapes, relative sizes, distances, and orientations of objects.

One area of interest is in whether the mental motion pictures system can represent language about "intangible" things which are important actors in texts about physical, biological, physical, and geological processes. As we build the conceptual lexicon to give our conceptual analyzer the ability to process a broader cross-section of texts in the ProPara dataset, we will have opportunities to explore how the mental motion pictures representation will be able to represent, for example, heat and light, and their creation, emission, and absorption events in the same way that it represents "solid" objects like "rain". Finally, many real world processes described in ProPara are cyclical, repetitive, or observed to occur as an accumulation of a particular event occurring repeatedly over time (e.g. the lunar cycle, erosion). Future explorations of how mental motion pictures and conceptual analyses may work for descriptions of these kinds of processes may benefit significantly from results appearing in the qualitative process theory literature (e.g. Forbus, 1984).

9. Acknowledgements

The authors thank Chris Riesbeck, Larry Birnbaum, and especially Mark Burstein for discussions which helped to shape this work and for publishing the ELI and Micro ELI code to make it available for our research efforts.

References

Lawrence Birnbaum and Mallory Selfridge. Micro ELI. In Roger C Schank and Christopher K Riesbeck, editors, *Inside Computer Understanding: Five Programs Plus Miniatures*, pages 354–372. Lawrence Erlbaum Associates, Hillsdale, NJ, 1981.

Anthony G Cohn, Brandon Bennett, John Gooday, and Nicholas Mark Gotts. Qualitative spatial representation and reasoning with the region connection calculus. *GeoInformatica*, 1(3):275–316, 1997.

Bhavana Dalvi, Lifu Huang, Niket Tandon, Wen-tau Yih, and Peter Clark. Tracking state changes in procedural text: a challenge dataset and models for process paragraph comprehension. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1595–1604, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

Michael G Dyer. *In-Depth Understanding: a Computer Model of Integrated Processing for Narrative Comprehension.* MIT Press, Cambridge, MA, 1982.

William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39, 2022.

K. D. Forbus. Qualitative process theory. *Artificial Intelligence*, 24:85–168, 1984.

Dagmar Gromann and Jamie C. Macbeth. Crowdsourcing image schemas. In *Proceedings of The Fourth Image Schema Day (ISD4)*, Bolzano-Bozen, Italy, December 2018. The International Association for Ontology and its Applications.

Lawrence Hunter, Zhiyong Lu, James Firby, William A Baumgartner, Helen L Johnson, Philip V Ogren, and K Bretonnel Cohen. OpenDMAP: an open source, ontology-driven concept analysis engine, with applications to capturing knowledge regarding protein transport, protein interactions and cell-type-specific gene expression. *BMC Bioinformatics*, 9 (1):1–11, 2008.

Ray S Jackendoff. *Semantics and Cognition*. MIT Press, Cambridge, MA, 1983.

Jamie C Macbeth. Enhancing learning with primitive-decomposed cognitive representations. In *International Workshop on Self-Supervised Learning*, pages 89–98. PMLR, 2020.

Jamie C. Macbeth and Dagmar Gromann. Towards modeling conceptual dependency primitives with image schema logic. In *The Fourth Workshop on Cognition And OntologieS (CAOS IV) at The Fifth Joint Ontology Workshop (JOWO'19)*, Graz, Austria, September 2019. The International Association for Ontology and its Applications.

Jamie C. Macbeth, Dagmar Gromann, and Maria M. Hedblom. Image schemas and conceptual dependency primitives: a comparison. In *Proceedings of The Joint Ontology Workshops, Episode 3: The Tyrolean Autumn of Ontology*, Bolzano-Bozen, Italy, September 2017. The International Association for Ontology and its Applications.

Samyam Rajbhandari, Jeff Rasley, Olatunji Ruwase, and Yuxiong He. Zero: Memory optimizations toward training trillion parameter models. In *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–16, 2020. doi: 10.1109/SC41405.2020.00024.

Roger C Schank. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, 3(4):552–631, 1972.

Roger C Schank. *Conceptual Information Processing*. Elsevier, New York, NY, 1975.

Aarohi Srivastava et al. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models, 2022. URL https://arxiv.org/abs/2206.04615.

Kristinn R Thórisson, Nivel Nivel, Bas R Steunebrink, Helgi P Helgason, Giovanni Pezzulo, Ricardo Sanz Bravo, Jürgen Schmidhuber, Haris Dindo, Manuel Rodríguez Hernández, Antonio Chella, et al. Autonomous acquisition of natural situated communication. *IADIS International Journal on Computer Science And Information Systems*, 9(2):115–131, 2014.

Anna Wierzbicka. *Semantics: Primes and Universals*. Oxford University Press, New York, 1996.

Yorick Wilks and Dann Fass. The preference semantics family. *Computers & Mathematics with Applications*, 23(2-5):205–221, 1992.

Zhuoran Zeng and Ernest Davis. Physical reasoning in an open world. *Proceedings of the Ninth Annual Conference on Advances in Cognitive Systems*, November 2021.

Wei Emma Zhang, Quan Z Sheng, Ahoud Alhazmi, and Chenliang Li. Adversarial attacks on deep-learning models in natural language processing: A survey. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(3):1–41, 2020.