Quarterly Interim Research Performance Report-
July – September 2013

C3: The Compositional Construction of Context

'A more effective and efficient way to marshal inferences from background knowledge'

N00014-13-1-0228

Dr. David McDonald
Smart Information Flow Technologies, dba SIFT, LLC
dmcdonald@sift.net

with Prof. James Pustejovsky
Brandeis University
jamesp@cs.brandeis.edu

1. What we've done

Presentations and publications In July, James gave a full day tutorial entitled *Actions, Objects, and their Habitats* at the Summer School in Cognitive Linguistics in Bangor, Wales, with this abstract.

> In this talk, I outline the challenges facing linguistic theories and computational language modeling in the attempt to contextualize linguistic knowledge in a non-linguistic world. I will focus on two aspects of contextualization of meaning: sentential semantic enrichment, involving extra-compositional mechanisms, such as coercion, co-composition, and other semantic devices used in Generative Lexicon Theory. The second aspect involves the embedding of linguistic utterances within the contexts of their use by speakers, mediated through deeper representational structures. I introduce the notion of "conceptual habitat", for the frame, script, and affordance based information associated with linguistic expressions. These habitats compose to create "minimal simulations", which are temporally traced minimal models, constructed from the use of linguistic expressions in specific contextualized frames, interpreted from a particular point of view. Minimal simulation is shown to be a computationally tractable approach to the dynamic interpretation of contexts, by virtue of the perspective-based constraints imposed.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
SEP 2013		**00-07-2013 to 00-09-2013**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
C3: The Compositional Construction of Context. 'A more effective and efficient way to marshal inferences from background knowledge'	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Smart Information Flow Technologies, dba SIFT, LLC,114 Waltham St Suite 24,Lexington,MA,02421**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**4**	

In September we jointly submitted a long paper to the December Advances in Cognitive Systems meeting: *On the Representation of Inferences and their Lexicalizations*. This paper drew substantially from our proposal and used its motivating example of how knowledge of the habitat and scripts associated with air travel set up the context for inferring who was stranded by canceled flights. It added our latest thinking on the computational architecture that makes this possible and illustrated the thin thread of implementation with the ISR example we presented at the June review meeting.

> Abstract: We have recently begun a project to develop a more effective and efficient way to marshal inferences from background knowledge to facilitate deep natural language understanding. The meaning of a word is taken to be the predications, entities, and potential inferences that it adds to an ongoing situation. As words compose, the minimal model in the situation evolves to limit and direct inference. At this point we have developed our computational architecture and implemented a thin thread through it on real text. Our focus is on efficient, fast execution and proving the feasibility of our design.

Also in September, James presented the paper *Dynamic Event Structure and Habitat Theory* at the International Conference on Generative Approaches to the Lexicon in Pisa Italy (`http://glcon2013.org`). Abstract:

> Abstract: In this brief note, I explore the cognitive mechanisms involved in interpreting the meanings of events, as conveyed through language. Specifically, I examine the notion of event simulation in the construction of linguistic meaning. Simulations are a special class of minimal models, generated from linguistic input, under a number of agent-oriented cognitive constraints. An integral part of this model is a dynamic representation of processes and events, such as the Dynamic Event Structure presented here. I show how simulations are composed of entity and event habitats, which are contextualization functions, acting to embed a proposition into a minimal model.

People At Brandeis we added a second graduate student to the project, Nikhil Krishnaswami, who will be more theory focused and complement the architecture and implementation focus of our other graduate student Alex Plotnick. At SIFT we continue to collaborate with Scott Friedman, particularly on questions of representation.

Research After the review meeting, our effort during the summer was dominated by working out how we design and develop the vision we had presented in June. Our example from the meeting encapsulated our task in a strikingly short amount of text: Given the context provided by the situation that is created by the sentence *Black Ford SUV has entered Wakil*, how do we make the inference from *Two people are dismounting* that these people were previously in the SUV?

The second sentence is picking out one particular aspect of the space of cohesive links and potential inferences provided by the focused object of the first sentence (the SUV). The connection is driven by the action *dismount*, but how did we know the connection was possible? Chuck Rieger had the same issues in his 1974 dissertation for Roger Schank, and drew this conclusion.

> "The fundamentals of understanding a story are rooted in the spontaneous expansion in a "multi-dimensional" inference space of each new thought as it arrives"

But we all knew, as did Rieger, that this unlimited chain of forward inference is intractable. The second sentence could have expanded on any aspect of the SUV (doors, windows, tailgate, driver, further motion,

. . .) or of the secondary theme of the village it just entered (*Three people are approaching*). To model the pycholinguistic facts, we wanted the access speed of a fully forward expanded set of possibilities, but without the computational cost of creating them or needing to know how to cut the inferencing off. We want the follow-on utterance to determine what is actually expanded.

Our working draft of a solution assumes that inferences ('the source of the people was the SUV) are going to come down to binding and unifying variables (source = 'people get out of the SUV'), while relying on the glue provided by the variables' lexical associations. We get the potential of a forward expansion by using a programming trick that lets us instantiate an arbitrary number of variables in unit time, i.e. to take knowledge of the form 'this concept licenses this variable,' and convert it to 'this particular individual (the SUV) has this variable.' That lets a trivial lookup of location properties in the situation take us to the SUV in one step, which mirrors the unconscious ease of human comprehension

2. What we're planning to do

The next steps of our research will be to flesh out our implementation and establish whether the scheme we managed to implement in just the one case will hold up when we enlarge the corpus it works from. There were several alternative 'pathways' we considered between *SUV* and *dismount*, and one of those might turn out to be superior.

Corpus The domain we proposed is modeling understanding route descriptions (hiking guides, travel blogs, etc.), which focuses on spatial concepts of location, movement and change in general. We have concluded that the simplest next step is to take the ISR[1] chat corpus that our example came from and handle all of it (two hours of observations over each of five days).

The corpus is essentially about the movement of cars and people along roads between villages. It has a small set of events and an established vocabulary, allowing us to model it thoroughly with just the limited resources we have. And it has the experimental advantage that we can test our implementation is a natural way by examining and modeling just a single day at a time, then testing how well the model performs on the next day's chat.

Ontology Up to this point we have selected our conceptual categories ad-hoc just by drawing on our experience. Now we will begin to a systematic and deliberate investigation of established upper and middle ontologies to determine what will provide us the best fit for our conception of lexicalized pragmatics. What factoring of the space of properties and implicatures is the most natural fit to the way language (English) cuts up knowledg?. What categories are the best for positioning a particular inference, default or affordance. What are the easiest to refactor and reverse?

Anticipated writing Within the research group we are going to make sense of our options and possibilities through a series of squibs. When they are mature we will pass them around to others for comments and critiques. We will draw on them to write full papers when good opportunities arise. In particular, we anticipate submitting a paper to next years International Generation Conference in June on the question of how a speaker is aware of the inferences that are likely to be available to their audience. We have found this dual perspective to be invaluable in the past.

Publication References

Pustejovsky, James P. (2013), *Actions, Objects, and their Habitats*, Summer School in Cognitive Linguistics, Prifysgol Bangor University, Bangor, Wales, July 24, 185 slides.

1. Intelligence, Surveillance, and Reconnaissance

McDonald, David D. & Pustejovsky, James P. *On the Representation of Inferences and their Lexicalizations*, submitted to the 2013 Advances in Cognitive Systems conference.

Pustejovsky, James P. (2013) *Dynamic Event Structure and Habitat Theory*, 6th International Conference on Generative Approaches to the Lexicon, September 24-25, 2013, Pisa Italy.