

Cognitive Modeling for Formal Semantics: The Organization of Discourse Representation Structures in Declarative Memory

Introduction. One of the key and still outstanding challenges for formal semantics is “how to build formal semantics into real-time processing models – whether psychological or computational – that involve the integration of linguistic and not-specifically linguistic knowledge.” (Partee 2011, p. 4) In this paper, we will outline the structure of a cognitively realistic semantic processor (already implemented), that is, a basic parser able to incrementally construct semantic representations in response to linguistic stimuli of the kind presented in self-paced reading or eye-tracking experiments.

Our choice for a processing-friendly semantics framework is Discourse Representation Theory (DRT, Kamp 1981; Kamp and Reyle 1993). We chose DRT because atomic DRSs and the compositional construction principles used to build them provide meaning representations and elementary compositional operations that are well understood mathematically, widely used in formal semantics, and can simultaneously function as both meaning representations / logical forms and their content / world knowledge / models (at least when we restrict ourselves to persistent DRSs; Kamp and Reyle 1993, pp. 96-97). Because of this double function, DRT and atomic DRSs can be thought of as mental models in the sense of Johnson-Laird (1983, 2004).

We also show how this semantic parser can be used to account for the reaction time (RT) data in the basic fan experiment reported in Anderson (1974). This fan experiment investigates how basic propositional information of the kind encoded by atomic DRSs is (stored and) retrieved from declarative memory, which is an essential component of real-time semantic interpretation in at least two respects: (*i*) incrementally processing semantic representations involves composing / integrating semantic representations introduced by new sentences or new parts of a sentence with semantic representations of the previous discourse; (*ii*) incremental interpretation also involves evaluating new semantic representations relative to our mental model of the world, and integrating their content into our world knowledge database stored in declarative memory. We provide the first (to our knowledge) cognitively realistic model of incremental interpretation and truth evaluation that is systematically informed by formal semantics and can be used to model RT data.

The fan effect. The fan effect “refers to the phenomenon that, as participants study more facts about a particular concept, their time to retrieve a particular fact about that concept increases.” (Anderson and Reder 1999, 186). The original experiment in Anderson (1974) demonstrated the fan effect in recognition memory. Participants studied 26 facts about people being in various locations, 10 of which are exemplified in (1) below.

-
- | | | |
|------------------------------|------------------------------|-----------------------------|
| 1. a. A lawyer is in a cave. | b. A debutante is in a bank. | c. A doctor is in a bank. |
| d. A doctor is in a shop. | e. A captain is in a church. | f. A captain is in a park. |
| g. A fireman is in a park. | h. A hippie is in a park. | i. A hippie is in a church. |
| j. A hippie is in a town. | | |
-

In the training part of the experiment, participants committed 26 items of this kind to memory. In the test part of the experiment, participants were presented with a series of sentences, some of which they had studied in the training part (targets) and some of which were novel (foils). They had to recognize the targets and had to reject the foils, which were novel combinations of the same people and locations.

The 10 items in (1) form a minimal network of facts that instantiates the 9 experimental conditions in Anderson (1974). These conditions have to do with how many studied facts are connected to each type of person and location. Different person and location concepts fan into 1, 2 or 3 sentences / facts. For example, *lawyer* and *cave* have a fan of 1, *captain* and *bank* have a fan of 2, and *hippie* and *park* have a fan of 3. The mean RTs, measured in s, for target recognition and foil rejection in this fan experiment are provided in the tables below (from Anderson and Reder 1999, p. 187).

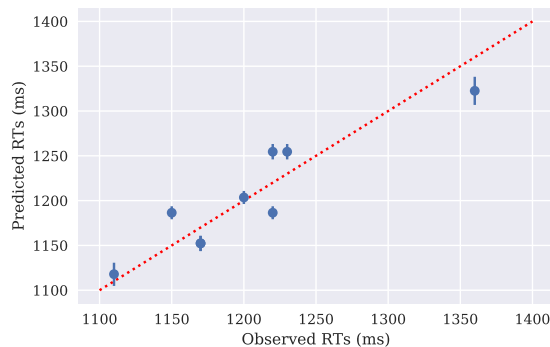
		Target RTs	location fan		
			1	2	3
person	fan	1	1.11	1.17	1.15
		2	1.17	1.20	1.23
		3	1.22	1.22	1.36

		Foil RTs	location fan		
			1	2	3
person	fan	1	1.20	1.25	1.26
		2	1.22	1.36	1.47
		3	1.26	1.29	1.47

Several **generalizations** become apparent based on this data: (i) averaging over targets and foils, the effect of 1-fan (both person and location) was about 1.2 s and increased by about 50 ms for each additional fan; (ii) the min effect: retrieval latency is a function of the minimum fan associated with a probe, e.g., participants tend to respond more slowly to the 2-2 fan items than to the 1-3 or 3-1 items; (iii) approx. equal fan effects for targets and foils, suggesting that foil rejection is not done by a serial (possibly exhaustive) search of the facts one knows about a cue / concept. Anderson and Reder (1999, pp. 188-189) account for these generalizations in ACT-R (Anderson 2007), a cognitive architecture that includes an explicit model of declarative memory, among other things.

The fan effect and the organization of DRSs in declarative memory. We can reformulate the notion of fan in Anderson’s experiment as a relation between the main DRS contributed by a sentence and the sub-DRSs contributed by its three parts: the person indefinite, the location indefinite, and the relational predicate *in*. Consider the fan example in (1j). The DRSs (meaning representations) of the three major components of the sentence – the indefinites *a hippie* and *a town* and the binary predicate *in* – are composed / combined together to form the DRS / meaning representation for the full sentence. This partitioning into 3 sub-DRSs matches the basic compositional skeleton generally assumed in the formal semantics literature for this type of sentences, as well as the real-time incremental comprehension process the ACT-R architecture imposes on us. The content of *a hippie is in a town* ends up represented in memory as the attribute-value matrix below.

MAIN-DRS	SUB-DRS ₁ :	DREF:	1
		PRED	hippie
		ARG1:	1
	SUB-DRS ₂ :	DREF:	2
		PRED	town
		ARG1:	2
	SUB-DRS ₃ :	PRED	in
		ARG1:	1
		ARG2:	2



The evaluation of the sentence as true (target) or false (foil) is accomplished by recalling from declarative memory the DRS with the highest activation and checking whether the currently built DRS matches its meaning representations (i.e., its sub-DRSs) or not. In principle, any DRS could be recalled from memory, but sub-DRSs in the currently built DRS spread activation to the same sub-DRSs in declarative memory, increasing the activation of the main DRS with matching information. Since memory search in ACT-R is not serial, we account for the generalization in (iii) above. Generalizations (i) and (ii) follow from properties of spreading activation (cf. Anderson and Lebiere 1998), and are discussed in the paper.

The paper also discusses how DRSs like the one above are built incrementally by means of production rules stored in procedural memory, and how the complete cognitive model is fit to the fan-experiment data by embedding it into a Bayesian model, for which it provides the likelihood function. The figure above plots the RTs predicted by our model against the observed, experimentally-obtained RTs (the blue dots), and 95% credible intervals around those predictions (the blue whiskers). The model captures the data well: the dots are close to the red diagonal line where predicted and observed RTs are identical. We end by discussing limitations of the model (the predictions are not exactly on the diagonal red line) and prospects for cognitive modeling in formal semantics.

References

- Anderson, John R. 2007. *How can the human mind occur in the physical universe?*. Oxford University Press.
- Anderson, John R., and Christian Lebiere. 1998. *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Anderson, John R., and Lynne M. Reder. 1999. The Fan Effect: New Results and New Theories. *Journal of Experimental Psychology: General* 128:186–197.
- Anderson, John Robert. 1974. Retrieval of propositional information from long-term memory. *Cognitive Psychology* 6:451 – 474.
- Johnson-Laird, Philip N. 1983. *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.
- Johnson-Laird, Philip N. 2004. The history of mental models. *Psychology of reasoning: Theoretical and historical perspectives* 179–212.
- Kamp, Hans. 1981. A theory of truth and semantic representation. In *Formal methods in the study of language*, ed. Jeroen Groenendijk, Theo Janssen, and Martin Stokhof, 277–322. Amsterdam: Mathematical Centre Tracts.
- Kamp, Hans, and Uwe Reyle. 1993. *From discourse to logic. introduction to model theoretic semantics of natural language, formal logic and Discourse Representation Theory*. Dordrecht: Kluwer.
- Partee, Barbara. 2011. The semantics adventure. URL https://udrive.oit.umass.edu/partee/Partee2011_MIT150.pdf, ms.