What is an action-based model of interpretation?

LAURA A. MICHAELIS

Scholars of linguistic meaning have always talked about sentences in terms of the things that they instruct their interpreters to do, find or check, and so the idea of an action-based model of interpretation is not entirely new. In denotational or 'realist' semantics, a sentence tells its interpreter how to determine its truth or falsity. In representational semantics, a sentence tells its interpreter to retrieve an array of semantic frames and make them fit together. When denotational semanticists started paying attention to narratives, they found an ingenious way to combine the representationalist idea of meaning as the integration of knowledge structures with the realist's view of meaning as truth conditions: they proposed that a sentence, rather than merely having a truth *value*, has a truth *interval* – the time for which it is true (Herweg 1991) – and that the interpreter can use linguistic and extralinguistic cues to determine whether or not two truth intervals overlap. Sometimes, however, the necessary contextual cues are lacking. For example, the English past tense, unlike its French analog, is aspectually neutral; therefore, the narrative passage in (1) is ambiguous:

(1) Esmé glanced into the orchestra pit. The conductor was furious.

This short text allows for two interpretations: one in which the conductor's fury was a result of Esmé's glance and another in which the furious state held prior to her glance and was merely witnessed by her. This ambiguity would not exist if the second sentence did not contain a state verb: only state sentences can be true at intervals greater than the intervals for which they are actually asserted to be true (Dowty 1986: 48). So, the problem of interpreting (1) comes down to that of properly constraining the truth interval of the state, and the only way to do that is to

Theoretical Linguistics 32–1 (2006), 65–71
DOI 10.1515/TL.2006.005
0301–4428/06/0032–0065

make a good guess. A narrative representation thus has a contingent quality: it is the best that someone could do under the circumstances, and it is subject to revision. The same can be said of plans, and it is the analogy between plans and discourse models that Hamm, Kamp and Van Lambalgen (henceforth HKL) exploit in their model of narrative understanding.

HKL argue that "the link between planning and linguistic processing is [...] provided by the notion of goal: we view a sentence S as a goal ('make S true') to be achieved by updating the discourse model". As HKL point out, a plan, like a discourse model, is a construction that contains a sequence of events and states and that takes into account both properties of the agent and properties of the world, like stable causal relationships. But for HKL, this analogy between plans and discourse models is only a point of departure: their work is a wholesale rethinking of narrative understanding. They argue that humans understand time in terms of contingency relations between states of affairs rather than in terms of a time line, and that narrative understanding would be impossible without the ability to compute plans to achieve goals. Their model of discourse processing builds on Discourse Representation Theory (DRT), a model-theoretic treatment of context modification developed by Kamp in the early 80s. HKL provide a computational implementation of DRT based on constraint logic programming. In this model, narrative inferences are represented as a process of query resolution. The queries are represented by formulas that contain operators (e.g., *Holds*, *Happens*, *Initiates*) which take situation types as their arguments. These arguments include fluents, *f*, events, *e*, and intervals, *t*. One such formula is used to represent the meaning of the English present-perfect construction (e.g., *The Eagle has landed*): *Initiates* (e, f, t). This means that an event initiated a state at some time. The present perfect fixes *t* as now, so query resolution amounts to determining whether the formula *Holds At* (f, now) can be added to the discourse model. In order to do this, the interpreter has to 'reason backwards' in the same way that planners go from a desired end state to the sequence of steps needed to bring about that end state. In the model of abductive reasoning proposed by HKL, the interpreter finds an axiom whose consequent clause is the same as a formula in the query, and then replaces the query formula with the antecedent of the axiom. The needed axiom is the one that HKL call the *axiom of temporal inertia*:

once initiated by an event, the consequent state of that event continues to hold unless terminated by a subsequent event. Since the interpreter knows that the fluent *f I have the flu* was initiated by an event *e*, and since no event terminating *f* has been mentioned, she or he can resolve the query concerning *Holds At* (f, now) by replacing the consequent of the temporal inertia axiom with its antecedent, *Happens* (e, t), t < now. The formula representing the initiating event therefore gets added to the model. Thus, making a present-perfect sentence true requires the interpreter to reason back to a causal event that happened before now. The moral of the story is that cause-effect sequences and our reasoning about cause-effect sequences run in opposite directions.

Because they view action planning and interpretation as the same cognitive faculty, HKL would find natural allies among proponents of simulation-based or embodied semantics, in which neurally inspired computational models of motor control are used to represent reasoning about actions (Feldman and Narayanan 2004, Narayanan 1999, Chang et al. 1998, Chang et al. 2003). In simulation semantics, actions are represented as a special type of Petri Net, called an executing- or x-schema. X-schemas are ordered collections of places and transitions (typed as instantaneous, hierarchical or durative transitions) that are connected by weighted directed arcs, typed as inhibitory, enabling and resource arcs. In this model, events and states are neither predications nor arguments of operators. Instead, a state is a place in the network and an event is a transition that connects an input place to an output place. X-schemas have a well defined run-time execution semantics that captures enablement, inhibition and resource consumption through a firing rule that specifies the circumstances under which a token in the network can move from one place to another. Because one schema can cause the continued execution of another schema, the model captures embedding relations, as in the schemas for *yanking*, *prying* and *plucking*, which involve simultaneous acts of grasping and pulling. The semantics of tense and aspect arise from the binding of verb-specific x-schemas to general x-schemas, called controllers, which capture general properties of complex events. Controller schemas map to grammatical markers of aspect and tense. Thus, while the meaning of the perfect construction is represented by a set of predications in the HKL model, in simulation semantics it is represented by the superimposition of one x-schema on another.

There is no question that the HKL model and embodied semantics share aims. Both are attempts to model the human capacity to reason about changing environments. Both are based on computationally implemented planning formalisms, and the axioms used for query resolution in the HKL model are implicitly encoded in the structure of x-schemas. So it is possible that the two models are just simulating the same inferences in different ways – when one model runs it crosses off predications, and when the other model runs it moves a token around. If the two models were merely notational variants of one another, the HKL formalism would be preferable, since it represents propositions about events in ways that are readily translatable into assertions, while the same cannot be said of x-schemas. But it precisely because embodied semantics is not based on operators and situation-type arguments that it appears to be a better model of linguistic meaning, in at least two respects. First, embodied semantics has fine-grained representations of verb meaning. Componential representation of verb meaning is important because a verb's syntactic behavior is a function of its aspectual class, as Rappaport Hovav and Levin 1998 show, and because some sentence meanings only make sense if modifiers can modify subparts of verbal Aktionsart representation, as in (2):

(2) She placed the bottle of champagne in the freezer for a few minutes.

The only reasonable reading of (2) is one in which the duration adverbial *for a few minutes* expresses the time that the champagne spent in the freezer. However, the availability of this reading can only be explained if the durational expression can select the resultant state entailed by the verb *place*. If, alternatively, such entailments were regarded as part of 'world knowledge', we would have to presume that a verb's syntactic behavior comes from pragmatics rather than from its more likely source, semantics. Second, embodied semantics uses the interaction between x-schemas and controller schemas to provide an explicit representation of the means by which grammatical constructions select components of lexical representations. For example, the French imperfective past tense selects for the class of state verbs (De Swart 1998). Indeed, while HKL make passing reference to Aktionsart-based notions like the telic-atelic distinction, their model of semantic type-shifting treats situation types as sets of eventualities rather than as complex networks of the type that have

to play a role in motor control. For this reason, the HKL model reveals very little about how humans represent actions and events. Such representations must minimally include the components that figure in lexical distinctions in the world's languages – parameters like intensity, direction, repetition, posture and volition. As Moens and Steedman (1988) showed, fine-grained event-structure representation provides an intuitive way to model the type shifts that grammatical constructions trigger. For example, the addition of a processual phase to a momentaneous transition yields a complex event representation that is then subject to stativization by means of the progressive construction, as in (3):

(3) Sue was fainting when Harry caught her.

As far as overlap relations are concerned, Sue's fainting is a state, but to describe the derivation of that state, one has to have a theory of what events can be embedded in what other events and how embedded transitions, states and processes may be selected by grammatical constructions like the progressive. Indeed, while Dowty (1986) presumes that inchoative verbs like *faint* cannot be combined with the progressive construction, Moens and Steedman correctly predict that they can, via coercion: in their approach to aspectual coercion, the verb *faint* may be augmented up to an accomplishment verb through the addition of a preparatory process (e.g., drooping), which in turn may be selected by the progressive, a stativizing construction which, according to Michaelis (2004), selects a stative 'rest' phase that lies between two type-identical events. Without fine-grained aspectual representations, new verb meanings cannot be built up from conventional ones, as in (3).

The necessity of embedding is shown by the wide array of contingency relations that can be signaled by *when*-clauses, as described by Moens and Steedman. They notice that the *when*-clause in (4) has a range of possible continuations, some of which are shown in (4a–c):

(4) When they built the 39th Street bridge,
  a. they used the best materials.
  b. a local architect drew up the plans.
  c. they solved most of their traffic problems.

In (4a), they argue, interpreting the relationship between main clause and subordinate clause requires selection of the preparatory process that led up to the creation of the bridge. In (4b), a preparatory process is invoked

that precedes the event of building. In (4c), the building event must be compressed into a point, and a consequent state added to it. The repertoire of possible type shifts is constrained because each type shift produces an event representation that either embeds or is embedded in the input event representations.

HKL make numerous references to the use of world knowledge in inferences about event relations, as when they claim that "[l]inguistic information must always be integrated with world knowledge". But why should world knowledge not *be* linguistic information? World knowledge informs action, and there is increasing evidence from neuroimaging studies that the action representations that we use to move around in the world are the same ones that we use for comprehending the actions of others:

> It appears that premotor and parietal areas are neurally integrated not only to control action, but also to serve the function of constructing an integrated representation of (a) actions together with (b) objects acted on and (c) locations toward which actions are directed. (Feldman and Narayanan 2004: 386)

According to proponents of simulation semantics, this integrated representation serves as the neural substrate of the meaning of action words. In other words, humans use their planning mechanisms for more than just updating discourse models – they use them to reason metaphorically about abstract events and actions in sentences like *The Indian government loosened its stranglehold on business* (Narayanan 1999), they use them for sentence comprehension (Glenberg and Kaschak 2002) and they use them to learn the meanings of action words from ostensive definition (Bailey 1997). While HKL have the laudable goal of connecting formal semantics to neuroscience, the planning formalism that they describe seems to owe more to formal semantics than it does to neuroscience.

Institute of Cognitive Science and Department of Linguistics
University of Colorado at Boulder

References

Bailey, David. (1997). When Push Comes to Shove: A Computational Model of the Role of Motor Control in the Acquisition of Action Verbs. Doctoral thesis, University of California, Berkeley, Computer Science Division.

Chang, Nancy, Daniel Gildea and Srinivas Narayanan. (1998). A Dynamic Model of Aspectual Composition. *Proceedings of the 20th Annual Conference of the Cognitive Science Society*. Madison, WI. 226–231.

Chang, Nancy, Laura Michaelis and Srinivas Narayanan. (2003). Together at Last: The Tense-Aspect Interface in Simulation Semantics. Paper presented at workshop on Embodied Construction Grammar, International Cognitive Linguistics Conference. University of La Rioja, Logroño, Spain.

De Swart, Henriette. (1998). Aspect Shift and Coercion. *Natural Language and Linguistic Theory* 16: 347–385.

Dowty, David. (1986). The Effects of Aspectual Class on the Temporal Structure of Discourse: Semantics or Pragmatics? *Linguistics and Philosophy* 9: 37–61.

Feldman, Jerome and Srinivas Narayanan. (2004). Embodied Meaning in a Neural Theory of Language. *Brain and Language* 89: 385–392.

Glenberg, Arthur and Michael Kaschak. (2002). Grounding Language in Action. *Psychonomic Bulletin and Review* 9: 558–565.

Herweg, Michael. (1991). A Critical Account of two Classical Approaches to Aspect. *Journal of Semantics* 8: 362–403.

Michaelis, Laura (2004). Type Shifting in Construction Grammar: An Integrated Approach to Aspectual Coercion. *Cognitive Linguistics* 15: 1–67.

Moens, Marc and Mark Steedman. (1988). Temporal Ontology and Temporal Reference. *Computational Linguistics* 14: 15–28.

Narayanan, Srinivas. (1997). *KARMA: Knowledge-based active Representations for Metaphor and Aspect*. Doctoral thesis, University of California, Berkeley, Computer Science Division.

—(1999). Reasoning about Actions in Narrative Understanding. *Proceedings of the International Joint Conference on Artificial Intelligence*. San Francisco: Morgan Kaufmann. 350–358.

Rappaport-Hovav, Malka and Beth Levin. (1998). Building Verb Meanings. In M. Butt and W. Geuder, eds., *The Projection of Argument Structure*. Stanford: CSLI Publications. 97–134.