1 On representing events – an introduction

Eric Pederson and Jürgen Bohnemeyer

This volume presents a collection of essays reporting on new research into the relationship between event representations in language and mind. In recent decades, linguists have increasingly invoked the notion of 'events' – under this and other labels – in modeling the meanings of natural language expressions. Indeed, numerous aspects of the structure of human languages are now commonly seen across theories and frameworks as geared towards the task of expressing event descriptions.

Like many of the constructs of semantic analysis and theory, the concept of 'event' has been influenced by the work of philosophers and natural scientists, usually with no more than a passing acknowledgment of the puzzles and controversies besetting its philosophical treatment (see Pianesi and Varzi 2000 for an overview). Philosophers have referenced the concept since antiquity, especially in treatments of causality (the subordinate notion of 'actions' has been used even longer in moral philosophy). However, events and their properties do not appear to have become topics of ontological research before the twentieth century, and their status must at present be considered far from settled. Even more glaring is the contrast between the rich and imposing architecture of event representations in language envisioned by many semanticists and the limited and scattered research on the status, nature, and role of event representations in the cognitive processing of perception and action by psychologists.

The research presented in this volume aims to make advances towards bridging the gap between linguistic and psychological research by illuminating from various perspectives the relationship between linguistic and cognitive event representations. The chapters come from different traditions and use different methods, but each presents empirical research on the interaction of linguistic and cognitive event representations. Some draw on data from the linguistic categorization of events in single languages (Pawley; Tversky *et al.*; Wolff *et al.*). Others directly compare results from multiple spoken (Bohnemeyer *et al.*; Slobin *et al.*; Carroll and von Stutterheim) or signed (Özyürek and Perniss) languages. Further, first language acquisition (Slobin *et al.*) and gestures accompanying speech (Gullberg) are examined. Attention and the visual

processing of stimuli during language production are examined (Dobel *et al.*). Two studies look at the non-linguistic categorization of event stimuli in the context of language use (the components of motion events in Loucks and Pederson; and event segmentation in Tversky *et al.*).

By presenting this set of different perspectives on the relationship between event encoding in language and internal cognition, the volume provides an overview of the research that has been conducted into this question. Our hope is that this will foster cross-stimulation, in that researchers interested in one approach (or method, or source of evidence) will find helpful the lessons from those pursuing other approaches.

1 Previous treatments of event representation in linguistics and psychology

Grammarians through the ages have relied on what one might think of as "expert folk theories" of event description in language. These are sets of unstated assumptions involving undefined notions that are presupposed by linguistic analyses. As an example, the practice of defining the verb as a part of speech or 'lexical category' (wholly or in part) with reference to the semantic property of describing (kinds of) actions or events can be traced back (in the European tradition) at least as far as the Greek grammarian Apollonius Dyscolus of the second century AD (Luhtala 2002: 279). Yet, explicit theories of event semantics would not be developed until the late twentieth century. It is impossible to characterize the assumptions folk theories consist of without turning them into something they are not – explicit statements. That said, the following core assumptions, even though they are couched in the terminology of contemporary linguistic theory, seem compatible with a great many of the folk theories implicit across the scholarship on language structure.

- Verbs generally describe (kinds of) actions or events.
- The arguments and complements of verbs – for example, subject, object, and perhaps certain kinds of embedded clauses – describe entities (or perhaps other events) involved in the event which is described by the verb (event participants).
- The roles that characterize the ways in which the participants are involved in the event – roles such as agent, theme, and recipient – are typically reflected by the syntactic properties of the expressions describing them. That is to say, the relationship between a verb and its arguments reflects these relationships between the event and its participants.
- The meanings of sentences and clauses involve states of affairs or propositions which may be about the reality or realization of the event described by the main verb of the sentence or clause.

To make this a little more concrete, consider the example in (1):

(1) Sally gave Floyd a book on event semantics on Monday with a conspiratorial wink

On the standard view of event encoding in contemporary linguistic theory, this sentence asserts a proposition concerning the occurrence of an event of the kind described by *give*, with the participant named by the subject, *Sally*, as the agent (here: the giver), the one named by the first (or 'primary') object, *Floyd*, as the recipient, and a third participant described by the second (and 'secondary') object, the noun phrase *a book on event semantics*, in the role of theme. All of the semantic properties just mentioned have been the focus of theorists' attention since the 1960s; but all of them have been part of implicit assumptions about event description from the beginning of scholarly work on the structure of language.

Indisputably, the most influential step in the development of event semantics was the publication of the paper 'The logical form of action sentences' by the philosopher Donald Davidson in 1967. Davidson's point of departure is a subtle observation: many adverbials, rather than functioning as true predicate modifiers, show an "intersective" behavior vis-à-vis the verb. For example, *on Monday* and *with a conspiratorial wink* in (1) do not so much single out particular kinds of giving, but rather impose independent constraints on the action described by the verb: the verb and its arguments require the action to be a giving of a book to Floyd by Sally, and the adverbials require the action to have taken place on a Monday and to have been conducted with a conspiratorial wink. In predicate-logic terms, it seems that the verb and the adverbials are all interpreted as predicates over the same argument, and that argument is not expressed by any of the syntactic arguments of the verb, but rather refers to the event itself. To formalize this insight, Davidson proposes that content words such as verbs and adverbs – and the nouns, adjectives, prepositions, and so forth that combine with them – express predicates, not just over individual arguments of the traditional kind referring to animate beings, inanimate things, and perhaps also abstract things, but over event arguments – existentially bound argument variables whose values are events.

Since 1967, Davidson's framework (and its numerous variants and offshoots) has been applied to many other problems of event semantics. One example is the theory of semantic (or 'thematic') roles such as agent, patient, recipient, and so on, alluded to above, the origins of which can be traced as far back as the Sanskrit grammarian Pāṇini of the fourth century BC. Most syntactic arguments of a sentence have referents that bear semantic roles in the event described by the sentence, and the syntactic properties of each argument reflect the semantic role of its referent. For example, in an English sentence with three syntactic arguments in the active voice such as (1), the agent is expressed by the

syntactic subject (*Sally*), the recipient by the first or 'primary' object (*Floyd*), and the theme by the 'secondary' object (*a book on event semantics*). Semantic roles mediate between the structure of an event and the syntactic structure of a sentence describing it, making the latter an abstract iconic representation of the former. This makes semantic roles key elements of what is often called the 'interface' between syntax and semantics, i.e., the principles that govern the mapping between form and meaning in language.[1]

In the 1970s, semanticists started thinking about how to model meanings that transcend the sentence level. The result was the family of so-called 'dynamic' approaches to semantics, which view meanings as properties, not of sentences, but of utterances in contexts. (Simplifying somewhat, sentences are complex linguistic signs composed of words and phrases, whereas utterances are actions that involve the use of such signs.) The most widely used of these frameworks is discourse representation theory (DRT; e.g., Kamp and Reyle 1993). From its beginnings, the modeling of the semantics of temporal operators such as tenses and viewpoint aspects has been one of the central goals of DRT. Viewpoint-aspectual meanings are illustrated by the contrast between progressive forms such as *was pushing* and simple past tense forms such as *drew*: the former present the events described in the sentences as ongoing at some reference point understood in context, whereas the latter describe these events as completed within a reference time frame. The miniature narrative fragment (2) is interpreted to the effect that the time during which Floyd's pushing is described as ongoing is identical to that during which Sally's drawing a circle is completed:

(2) Floyd was pushing a cart. Sally drew a circle

In DRT, the sentences in (2) are modeled as introducing the 'run times' of the events of Floyd pushing a cart and Sally drawing a circle as 'discourse referents' with the tenses and viewpoint aspects encoding relations between these times. An even more radical departure from the Aristotle–Frege tradition in logical semantics is proposed in another dynamic theory of linguistic meaning, situation semantics (Barwise and Perry 1983). Instead of expressing, or being about, propositions, which are true in certain possible worlds and false in others (i.e., are functions from possible worlds to truth values), the meanings of utterances are assumed in situation semantics to be exemplified by situations which 'support' or 'exemplify' the utterances. The notion of 'situation' is given wide interpretation in this approach, ranging all the way from Davidsonian events on one end to possible worlds (Lewis 1986) on the other. Among approaches to semantics that view meanings in terms of relations between utterances and mental states or cognitive representations, Jackendoff's

[1] The theory of semantic roles has been formalized in a Davidsonian framework in Carlson (1984) and Parsons (1990).

(1983) conceptual semantics takes a position similar to that of situation semantics, viewing the highest-order conceptual functions expressed in sentences and utterances as characterizing events and states rather than propositions. These few remarks should convey how explicit event semantics has gradually developed over the course of the past decades into something approaching a common metalanguage which can be used by researchers of diverse theoretical backgrounds in dealing with a vast array of aspects of the structure of natural languages.[2]

In psychology, the question of the relationship between object perception and event perception has played a significant role in perception research. James J. Gibson and Gunnar Johansson independently of one another amassed evidence to the effect that not only are there distinctive gestalt patterns of events in the flow of perceptual information (Gibson's 'styles of change,' von Fieandt and Gibson 1959), but event perception must in fact at least in some ways be prior to object perception. Gibson's and Johansson's (1973, 1975) approaches to perception differ in the amount of inherent mediating information they ascribe to the perceptual system. Johansson assumes that the perceptual system operates on certain rules that allow it to prioritize in cases of informational ambiguity in the input, rules that, e.g., favor a rigid-rotation perception over an elastic-deformation perception in case the input permits both construals. Gibson, in contrast, proposes a theory of 'direct perception' in which the perceptual system is assumed to be attuned to an environment informationally rich enough so as not normally to give rise to the kinds of ambiguities that occur under lab conditions. This approach has become known as the 'ecological' approach to perception psychology. Gibson attempts to minimize the role of event concepts or 'schemas,' arguing that there is a principled coordination of the schematic properties of perceived events and those of actions potentially carried out by the observer: the schematization of objects and events is based on 'affordances' of the objects and events, i.e., those properties that are functionally relevant for potential action in the environment. Surveys of the early work on event perception and cognition in the Gibsonian framework are offered by Warren and Shaw (1985) and McCabe and Balzano (1986). The idea of direct coordination between event perception and action has more recently been taken up in Prinz's common coding theory (e.g., Prinz 1997). The three-volume *Handbook of Perception and Action* edited by Prinz and colleages (Prinz and Bridgeman 1995; Heuer and Keele 1996; Neumann and Sanders 1996) provides a comprehensive overview with a focus on the conditions for an integrated account of perception and action.

[2] For more comprehensive overviews of the history of event semantics, see Higginbotham (2000) and Rothstein (1998b). Tenny and Pustejovsky (2000) survey the role of event semantics in the theory of language structure.

Like Gibson's and Johansson's work, the ground-breaking work conducted by A. Michotte and his collaborators on the processing of causal information in perception was influenced by gestalt psychology. In a classical study described in Michotte and Thinès (1963), participants were shown displays of virtual ballistic collisions between two rectangles. The participants were aware that the rectangles were not actually three-dimensional objects and their movements were not actually ballistic. They nevertheless reported seeing a variety of different types of ballistic collisions, the precise type depending on parameters such as the length of the time interval between the contact of the two rectangles and the beginning of the motion of the second rectangle.

For more recent replications of some of Michotte's studies with contemporary stimuli and response measures, see Schlottman and Anderson (1993). Michotte interpreted his findings to the effect that participants directly perceive the causality in these events, rather than merely inferring it from the succession of sub-events, as Hume's (1739–1740) classical account of causal reasoning suggests. Verfaillie and Daems (1996) used Michotte-style ballistic collision stimuli in a chronometric study in which participants identified agents and patients. This study provided evidence for a cognitive basis of the theory of semantic roles mentioned above.

Newtson (1973), working within the field of social psychology, conducted classic studies on the role of mereological (part–whole) structures in the cognitive encoding of events. Newtson showed participants a video in which an actor filled in a questionnaire, smoked a cigarette, and read in a book. The participants were then asked to segment the events in the clip into units by pressing a button at event boundaries ('break points'). Half of the participants were given a task of fine-grained segmentation, the other half were asked to perform coarse-grained segmentation (Newtson 1973: 30). Correspondence of fine-grained and coarse-grained break points was found to be relatively low. This led Cohen and Ebbesen (1979) to question the existence of task-independent knowledge of event mereologies. However, Zacks and Tversky (2001) repeated Newtson's experiments with different stimulus scenes and more precise measures of event boundaries and found a much higher rate of coincidence across coarse-grained and fine-grained segmentations than Newtson (see also Tversky *et al.*, this volume). Other classic studies that have produced evidence of mereological knowledge in event cognition include Newtson and Engquist (1976) and Jenkins, Wald, and Pittenger (1986). Jenkins *et al.* (1986) prepared a series of picture stills representing three different event sequences. In one sequence, a woman prepares a cup of tea; in a second, a teenage girl is shown having a conversation on the telephone, and the third series shows stills from a party. Unlike the first two, the third series of pictures does not represent one coherent event sequence. Every third picture was removed from each sequence. The participants were then tested for recognition of pictures from a randomly ordered

subset of the pictures shown originally and foils from the subset that had been taken out. The authors demonstrated that the participants were more likely to detect foils from the less coherent event sequence. They argued that a less coherent sequence induces participants to memorize each individual picture, rather than to encode the entire series as a representation of an event. This suggests that event concepts may spell out mereological information and that dynamic stimuli that do not conform to any recognizable mereological schema are not readily conceptualized and memorized as instances of complex events.

It is against the backdrop of this scholarship that the chapters in the current volume contribute to the study of event representations in perception, action, mind, and language.

2 The current volume

Half of the chapters in this volume directly draw data from crosslinguistic comparison in their investigation of the relationships between form, meaning, and conceptual representation. By asking what properties of linguistic event representations vary across languages and what are universal, crosslinguistic approaches open one empirical window on the principles that govern the mapping between event representations in language and internal cognition.

Andrew Pawley kicks off this series of comparative chapters by sketching a somewhat extreme example of what is possible in terms of the lexical and syntactic resources used in event descriptions. Kalam, a language of the highlands of Papua New Guinea, has only a small set of around a hundred verb roots with very elementary meanings.[3] The vast majority of meanings lexicalized in simple verbs in English require various kinds of so-called 'serial verb constructions' for their expression in Kalam. Such constructions involve strings of two or more verbs which are not syntactically dependent upon one another, but may together form a single clause or phrase. Serial verb constructions not only render meanings expressed by single verbs in English, but also meanings expressed by combinations of verbs and prepositional phrases or verbs and embedded clauses.

In his classic 1987 paper, Pawley argued that at least for the purposes of talking about them, Kalam speakers must conceptualize events quite differently from how they are conceptualized by native speakers of English. This conclusion was challenged by Givón (1990, 1991a) on the basis of experimental data. Givón compared descriptions of a video clip by speakers of Kalam,

[3] This is not a unique case: languages with closed and very small classes of verb roots appear to be common both on New Guinea and in northern Australia (there are also reports from other parts of the world; e.g., Dickinson 2002 and Sakel 2004 describe two unrelated languages of the Andes with similar phenomena).

three other Papuan languages, and Tok Pisin, an English-based Creole spoken in Papua New Guinea and elsewhere in Melanesia. He found that serial verb constructions are not likely to be interrupted by pauses, suggesting that they are produced as chunks, like single words. Givón took this as evidence that the same cognitive event representations expressed by single-verb sentences in English can be expressed by serial verb constructions in languages such as Kalam.[4]

In his contribution to the present volume, Pawley provides his response. He draws a distinction between two types of serial verb constructions, which he terms 'compact' and 'narrative.' Compact series do not permit the insertion of any material between the verbs that constitute the series. Their meanings can generally be expressed by single verbs in English. In contrast, narrative series may span multiple clauses and correspond to multi-clausal narratives in English, following similar principles in terms of the order of presentation and the type of information presented at each stage. Yet, like compact series, they typically fall under a single intonation contour, and each sequence of verbs appears to be stored as a template in long-term memory. Narrative series thus provide evidence even by Givón's criteria that the segmentation of events in the preverbal message during narrative production differs between speakers of English and Kalam.

The Pawley–Givón debate illustrates several larger points which serve as a good introduction to this set of chapters. First, languages differ quite drastically in the lexical and syntactic resources they provide for event descriptions and the constraints they impose on the use of these. Second, research into the semantic impact of this variation – and thus into crosslinguistic variation and uniformity in the event representations expressed in language – has been rather preliminary to date and there is no agreement on methodological standards. Thirdly, attitudes towards the question of semantic variation have been heavily informed by the universalism–relativism and nature–nurture debates: what aspects of language and cognition are language-specific and/or culture-specific and thus presumably learned, and what universal and thus potentially innate? Attempts at shedding empirical light on these questions have all too often been overwhelmed by ideological preconceptions and prejudices.

The second of the above points has recently been addressed by Bohnemeyer *et al*. (2007). Typologists – students of language variation and universals – have long used an intuitive distinction between constructions that describe "single events" vs. "multiple events." For example, *Floyd went from Rochester to Buffalo* might be said to describe a single event, whereas *Floyd left Rochester and arrived in Buffalo* might be said to describe a sequence of two events. Bohnemeyer *et al*. (2007) propose that this intuition can be formalized using

[4] For a recent typological survey of serial verb constructions, see Aikhenvald and Dixon (2006).

the "scopal" properties of temporal operators as a criterion: descriptions that intuitively refer to a sequence of multiple events allow the speaker to "time" the sub-events independently of one another (e.g. *Floyd left Rochester at eight and arrived in Buffalo at nine*), whereas descriptions that intuitively refer to a single event do not (?*Floyd went from Rochester at eight to Buffalo at nine* sounds odd). They call this property of event descriptions being compatible only with those time expressions that refer to the time of the entire larger event the 'macro-event property' (MEP) and the descriptions that have the MEP 'macro-event expressions.' Equipped with the methodological innovation of the MEP, Bohnemeyer *et al*. (2007) find variation in the representations of motion events far exceeding the assumptions of previous work. But the study also uncovered principles that are shared across the languages of the sample. In their contribution to this volume, *Jürgen Bohnemeyer*, *N. J. Enfield*, *James Essegbey*, and *Sotaro Kita* present a case study that applies the methodology of Bohnemeyer *et al*. (2007) to a new domain: the expression of causality. They examine the segmentation of causal chains into macro-event expressions in four unrelated languages: Ewe, Japanese, Lao, and Yukatek Maya. Like the study on motion event segmentation before it, the present study shows that languages differ in the events for which they provide macro-event descriptions. The source of these differences is variation in both the availability of lexical expressions for concepts and syntactic constructions to combine these.

Mary Carroll and *Christiane von Stutterheim* examine the impact of language-specific patterns in the mapping between information perspective and syntax on event descriptions. Information perspective identifies referents as new to the discourse vs. already established, as foregrounded vs. backgrounded in an utterance, etc. English and German, though closely related, differ in the interface between information perspective and syntax. In English event descriptions, new referents are by preference introduced in existential predications, bumping the categorization of the event into a second clause which is often subordinate (e.g., *There is a girl shopping in a supermarket*). In contrast, German event descriptions freely permit introduction of new referents in indefinite noun phrases, thereby allowing the categorization of the event to take place in the same clause (e.g., 'A girl shops in a supermarket'). This is the first study ever to look at the role of information perspective in the structure of event descriptions. The authors present evidence of secondary effects on the segmentation of event descriptions.

Turning to event representation in the sign languages of the Deaf, *Aslı Özyürek* and *Pamela Perniss* look at Turkish and German Sign Language and the ways these are anchored in the established reference space and with respect to the signer's body. In order to represent events about action, motion, and location (e.g., to depict flipping a pancake), signers need to project the referents and the event space onto their body and the space around them. The authors

investigate the similarities and differences in perspective choice and its interaction with event descriptors in these sign languages. They suggest that although the visual-spatial modality might constrain and homogenize expressive possibilities in sign languages, there remains diversity in the expression of events across sign languages just as is reported for spoken languages. Better understanding this bodily and spatial linguistic expression of events can broaden our understanding of how events are represented in languages more generally.

The next two chapters focus on Talmy's (1985, 2000b) well-cited typology of verb-framed vs. satellite-framed motion descriptions. Verb-framed descriptions express the path component (information about from/to where the 'figure' moves) in the main verb root; satellite-framed descriptions express the path peripherally to the verb (e.g. adverbially or in a particle). Languages tend to systematically favor one type of description or the other, based on lexical and syntactic factors. Several studies in recent years have investigated the question whether speaking one type of language or the other (especially as one's native language) influences the cognitive processing of motion events. In their chapter, *Jeff Loucks* and *Eric Pederson* report on two studies they conducted with speakers of English, Japanese, and Spanish, involving the categorization of human motion events. A separate set of speakers were asked to describe these same motion stimuli. There appears to be no general support for cognitive effects of Talmy's patterns in that all groups demonstrated no consistent bias in their categorization strategies. Loucks and Pederson conclude with suggestions for revising Talmy's typology for these purposes as well as critiquing the methods so far employed by this line of research.

Dan I. Slobin, *Melissa Bowerman*, *Penelope Brown*, *Sonja Eisenbeiß*, and *Bhuvana Narasimhan* use child and adult language descriptions of placement ("putting") events in four satellite-framed (English, Finnish, German, Russian) and four verb-framed (Hindi, Spanish, Turkish, Tzeltal Maya) languages to examine the extent to which child language follows patterns of adult variation or is largely constrained by pre-linguistic and universal notions of event encoding. This is one of the first studies examining the developmental patterns of event encoding. While the languages can be roughly categorized following Talmy's typology, the authors find fine-grained crosslinguistic variation within each of the two groups. For those event features which are perceptually salient, even quite young children prove sensitive to these finer-grained adult-language characteristics. The authors argue that a "multiplicity of interacting factors . . . each with its own language-specific constraints and regularities" must be assumed to account for this variation in child sensitivity to input variation.

Also looking at the expression of placement events, *Marianne Gullberg* investigates co-speech gestures in Dutch and French. Taking gesture as indicative of at least some aspects of underlying cognitive event representations, her