The semantic categories of CUTTING and BREAKING events: A crosslinguistic perspective

ASIFA MAJID, MELISSA BOWERMAN, MIRIAM VAN STADEN and JAMES S. BOSTER*

Abstract

This special issue of Cognitive Linguistics explores the linguistic encoding of events of CUTTING *and* BREAKING. *In this article we first introduce the project on which it is based by motivating the selection of this conceptual domain, presenting the methods of data collection used by all the investigators, and characterizing the language sample. We then present a new approach to examining cross-linguistic similarities and differences in semantic categorization. Applying statistical modeling to the descriptions of* CUTTING *and* BREAKING *events elicited from speakers of all the languages, we show that although there is cross-linguistic variation in the number of distinctions made and in the placement of category boundaries, these differences take place within a strongly constrained semantic space: across languages, there is a surprising degree of consensus on the partitioning of events in this domain. In closing, we compare our statistical approach with more conventional semantic analyses, and show how an extensional semantic typological approach like the one illustrated here can help illuminate the intensional distinctions made by languages.*

Keywords: CUT *and* BREAK*; separation events; verb semantics; categorization; extension; intension; typology; semantic map.*

1. Introduction

To be able to talk about their experiences, speakers have to parse the ongoing perceptual flow into units and categorize these units as instances of recurrent event types associated with words of their language. Where do the event categories labeled by everyday verbs such as English *walk*, *come*, *go*, *put*, *eat*, *open*, and *hit* come from, and how widely are they shared across languages and cultures?

Cognitive Linguistics 18–2 (2007), 133–152
DOI 10.1515/COG.2007.005
0936–5907/07/0018–0133

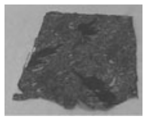
According to one view, the categories associated with everyday words are largely universal. This is because, by hypothesis, they originate in nonlinguistic cognition, and are shaped by perceptual and cognitive predispositions, environmental and biological constraints, and activities common to people everywhere. In this view, acquiring basic vocabulary is a process of mapping words to concepts that have already been established on a nonlinguistic basis (Bloom 2000; E. Clark 1973; H. Clark 1973; Gleitman 1990; Piaget 1954; Slobin 1973). A contrasting view holds that lexical categories do not reflect shared nonlinguistic cognition directly, but are to some extent linguistic conventions that are free to vary—no doubt within limits—according to historical, cultural, and environmental circumstances (Bowerman and Choi 2001; Malt et al. 1999, 2003; Wilkins and Hill 1995). According to this point of view, learning words, even for seemingly universal human experiences and activities, often involves working out language-specific principles of categorization (Bowerman and Choi 2001, 2003; de León 2001; Pye et al. 1995).

This special issue of *Cognitive Linguistics* examines cross-linguistic universality and variation in the encoding of a particular domain of events, those involving "separation in the material integrity of objects" (Hale and Keyser 1987)—CUTTING and BREAKING (or C&B) events. Although its boundaries are somewhat vague, the C&B domain is taken to centrally include at least the kinds of events known in English as cutting, breaking, slicing, chopping, hacking, tearing, ripping, smashing, shattering, snapping, and so on (see Levin 1993, §21 on verbs of CUTTING, and §45.1 on verbs of BREAKING).

This domain was chosen for several reasons. On the one hand, actions of C&B have been central to hominid cognition and culture for more than two million years (see Toth and Schick 1993, on the fossil tool record), which might plead for a degree of universality in the conceptualization of such events. Consistent with this, C&B verbs have figured prominently in recent discussions of universals of verb semantics and syntax: the underlying semantic structure of CUTTING-type verbs, it is claimed, is distinct from that of BREAKING-type verbs, and the two kinds of semantic structures are associated with distinct argument structure and syntactic privileges (Guerssel et al. 1985; Levin and Rappaport 1995; see Bohnemeyer, this issue).

On the other hand, preliminary cross-linguistic work (Pye 1994, 1996; Pye et al. 1995) shows that C&B verbs have intriguingly different extension patterns in different languages. This is illustrated in Table 1, adapted from Pye (1994), which shows the extensions of the words conventionally used to describe actions of "separation in the material integrity" of four

Table 1. *Comparison of C&B verbs in English, Garifuna, and Mandarin (Pye 1994)*

	 cloth	 bubble	 plate	 stick
English	*tear/rip*	*pop*	*break*	
Garifuna	*teiriguana*	*bowguana*		*halaguana*
Mandarin	*noŋ4-puo4*			*noŋ4-duan4*

objects—a piece of cloth, a bubble, a plate, and a stick—in English, Garifuna, and Mandarin.

The existence of cross-linguistic diversity in the partitioning of C&B events suggests that the necessary set of categories may not be obvious to first language learners purely on the basis of nonlinguistic experience, but must be learned through exposure to the input language. There is indeed evidence for such a learning process. Children make many errors in their spontaneous speech in referring to events of C&B, often applying a C&B word where an adult would not use it, or applying a word from another semantic class to a C&B event. Examples from young learners of English include *Daddy cut ice* (age 1;10—one year; 10 months), for breaking ice cubes into chips with a rolling pin, *Open. Cut.* (1;8), as the child pulls a grapefruit section apart with her fingers, *Don't break my coat* (2;11), as someone pulls on the child's coat, and *Open* (1;7) for breaking the leg off a plastic doll (Bowerman 2005). Analogous errors were made in an elicited production task by 3- to 5-year-old learners of Mandarin, K'iche' Mayan, and English (Pye et al. 1995). For example, up to 31 percent of 4-year-olds applied a core BREAKING verb of their language (Mandarin *duan4* 'be.broken.crosswise (of a long thing)'; K'iche' *q'upi:j* 'break a hard thing'; English *break*) to events of cutting paper with scissors.

Errors with C&B verb categories persist for many years. Even at age 7, learners of English judge that the verb *break* can be used for peeling an orange, *cut* for breaking a bottle by bashing it with a pair of scissors, and *open* for breaking a cracker with the hands or cutting an apple with a knife; no adult shares these judgments (Schaefer 1979). Schaefer suggests that learning the adult categories is a drawn-out process of isolating the features associated with the verbs and adjusting their weights appropriately. Consistent with such a learning process, patterns of overextension are language-specific rather than universal. That is, errors seem to be governed not by universal event concepts shared by all children, but by properties of the specific semantic categories to be learned, including category size, word frequency, and the number and nature of competing lexical categories (Bowerman and Choi 2003).

There are, then, arguments both for and against the hypothesis that core categories of C&B events are cognitively obvious and universally shared. This makes C&B an attractive focus for a large-scale cross-linguistic project on event encoding. Such a project was designed by members of the Event Representation Project at the Max Planck Institute (Nijmegen, The Netherlands), and carried out, together with colleagues elsewhere, in a variety of field sites around the world. Drawing on concepts and techniques from linguistics, psychology, and anthropology, the authors of the articles in this special issue of *Cognitive Linguistics* present their results on the extent and nature of cross-linguistic variation in the linguistic encoding of C&B events.

The issue is organized as follows. The remainder of this first article has two goals: (1) to describe the stimulus materials and methods of data collection used by all researchers in the Cut and Break project, as well as to introduce the sample of languages investigated, and (2) to explore the structure of the linguistic categorization of C&B events from a broad cross-linguistic perspective by analyzing data from the entire sample with multivariate statistical techniques. Bohnemeyer (this issue) overviews data from the whole sample as well, in this case with the relationship between verb semantics and syntax in mind. His goal is to test the hypothesis (Guerssel et al. 1985; Levin and Rappaport 1995) that across languages, the semantic structure of CUTTING-type verbs differs systematically from that of BREAKING-type verbs, with a concomitant difference in the kinds of argument structure alternations the two classes of verbs allow. Then follows a series of studies devoted to more detailed analysis of the linguistic representation of C&B events in a number of specific languages or language groups. The focus of these articles varies according to researchers' interests, but always includes the key features of C&B event encoding in the language(s) at issue, and especially any surprising or clearly language-specific features.

2. Collecting cross-linguistic data on the encoding of C&B events

2.1. *Developing an elicitation tool*

To establish an empirical data base for within- and across-language analysis, C&B project members created a set of videoclips depicting C&B events and related scenes, to be used in eliciting comparable event descriptions from speakers of diverse languages (Bohnemeyer et al. 2001). Not all conceivable C&B events could be depicted, of course, nor could speakers be expected to sit still for hundreds of clips. We tried nevertheless to sample well beyond the distinctions and groupings made by the C&B verbs of English and other familiar European languages. Here, the cross-linguistic work of Pye and his colleagues (Pye 1994, 1996; Pye et al. 1995) proved a valuable source of inspiration for candidate distinctions to test, as did the field experiences of project members and colleagues working on far-flung field languages.

It was also important not to restrict our sample of events according to our preconceived ideas of what constitute events of C&B, since languages might have forms that apply not only to events that English speakers would call e.g., cutting, breaking, or smashing, but also to events that they would call e.g., *taking apart* or *opening*. Here, attested overextensions of C&B verbs in children's spontaneous or elicited speech provided useful guidelines, since children's errors within any one language suggest possible dimensions along which human cognizers—hence languages—might compute similarities among events. The kinds of events to which children often overextend C&B verbs—or from which they borrow verbs to describe C&B events—are events that involve separation but with no or minimal material destruction (they are often reversible): these include, most prominently, events of opening, pulling apart, pushing apart, taking apart, and peeling (Bowerman 2005; Schaefer 1979).

The final C&B elicitation tool consists of 61 videoclips depicting separations of various kinds; a short description is provided in the Appendix.[1] Most clips show separations with material destruction (i.e., seemingly core events of C&B). The majority of these include a causal agent, but a few depicted a seemingly spontaneous separation (e.g., a twig snapping). The physical properties of the affected objects were varied (e.g., stick, rope, cloth, plate, pot, hair, food items), as were the instruments (e.g., hand, knife, scissors, karate-chop, machete, hammer) and the manner of the action (e.g., once or repeatedly, calmly or with furious intensity). In addition to core C&B events, the set of videoclips also included separations such as opening a teapot, the hand, or a book; taking the top off a pen; pulling apart paper cups; and peeling a banana.

2.2. *Language sample*

Using the C&B elicitation tool, 24 researchers collected event descriptions from a total of 28 languages, as listed in Table 2. For each language there was an average of three speakers (range 1;0 to 7;0). Although these languages constitute a convenience sample based primarily on the

(AutoPDF V7 15/6/07 10:43) WDG (148×225mm) TimesM J-1787 COGL, 18:2 HC1: WSL(A) 14/06/2007 pp. 133–152 1787_18-2_01 (p. 137)

Table 2. *The 28 languages from which C&B data were collected*

Language family	Language	Researcher
Altaic	Turkish	A. Özyürek
Austronesian	Biak	W. van de Heuvel
	Kilivila	G. Senft
Cariban	Tiriyó	S. Meira
Dravidian	Tamil	B. Narasimhan
Indo-European	Dutch	M. van Staden
	English	M. Bowerman, A. Majid
	German	M. van Staden
	Hindi	B. Narasimhan
	Punjabi	A. Majid
	Spanish	E. Palancar, M. Bowerman
	Swedish	M. Gullberg
Mayan	Tzeltal	P. Brown
	Yukatek	J. Bohnemeyer
Niger-Congo	Ewe	F. Ameka, J. Essegbey
	Jalonke	F. Lüpke
	Likpe	F. Ameka
Otomanguean	Otomi	E. Palancar
Pama-Nyungan	Kuuk Thaayorre	A. Gaby
Sino-Tibetan	Mandarin	J. Chen
Tai	Lao	N. Enfield
West Papuan Phylum	Tidore	M. van Staden
Witotoan	Miraña	F. Seifart
Creole	Sranan	J. Essegbey
Isolate	Chontal	L. O'Connor
	Japanese	S. Kita
Isolate (Papuan)	Touo	M. Dunn, A. Terrill
	Yélî Dyne	S. Levinson

researchers' field languages, they are nonetheless typologically, genetically, and geographically diverse, representing 13 different language families, along with four isolates and a Creole. Indo-European languages are overrepresented, which might lead to an overestimation of overall language similarity since related languages often share features (cf. Dryer 1989; Perkins 1989; Rijkoff and Bakker 1998). But the danger of overestimating language similarity is lower for features of language that change rapidly than for those that are conservative (Rijkoff and Bakker 1998). Relative to grammatical features, lexical semantic categorization may change rather quickly (e.g., Ross 1996). There are, for example, major differences in how cognate prepositions of English, German, and Dutch categorize topological spatial relations (cf. Bowerman 1989, 1996), and a comparison of the four Germanic languages in our sample also reveals striking differences in the categorization of C&B events (Majid et al., this issue).

(AutoPDF V7 15/6/07 10:43) WDG (148×225mm) TimesM J-1787 COGL, 18:2 HC1: WSL(A) 14/06/2007 pp. 133–152 1787_18-2_01 (p. 138)

2.3. *Procedure*

Consultants were shown the videoclips in a fixed order. After viewing each clip, they described the event they had seen. Further linguistic probing, e.g., for alternative descriptions, was carried out at the discretion of the researcher. The descriptions were tape- or video-recorded and later transcribed and coded. The researchers who collected the data are experts in the languages they investigated. This means that they could instruct and interact with consultants directly in the target language rather than in a contact language that might influence the event descriptions obtained. Their expertise was also critical for coding and analyzing the data and interpreting the results, since their knowledge of the lexical items and constructions elicited from speakers of "their" language was based on language materials extending far beyond this project.

3. Cross-linguistic comparison of the linguistic categorization of C&B events

Our description of the C&B project up to this point serves as background to all the articles in this special issue. With this stage-setting complete, we now turn to our own study, a large-scale statistical analysis that draws together data from all 28 languages in the sample to explore cross-linguistic similarities and differences in the semantic categorization of C&B events.

3.1. *Units of analysis*

An important prerequisite to linguistic comparison is to select the elements to compare. In our study this requires some thought, since, as is well known, languages differ in their characteristic way of distributing information over clause constituents (Talmy 1985, 1991). The critical information in the C&B project concerns the linguistic categorization of a state change involving some kind of separation in an object (usually explicitly caused by an agent, sometimes not). State change is a "framing event", according to Talmy (1991); the "upshot" or core of what the sentence asserts.

In some languages (e.g., Yélî Dyne in Levinson, this issue), speakers routinely encoded information about the (caused) state change shown in the videoclips in a single monomorphemic verb; this verb is then obviously our unit of analysis. In other languages, (e.g., Mandarin in Chen, this issue) speakers typically distributed the information across two (occasionally more) component verbs of a compound verb: e.g., *bai1-duan4* 'snap-be.broken.crosswise (of a long thing)' and *jian3-kai1* 'cut.with

(AutoPDF V7 15/6/07 10:43) WDG (148×225mm) TimesM J-1787 COGL, 18:2 HC1: WSL(A) 14/06/2007 pp. 133–152 1787_18-2_01 (p. 139)

.scissors-be.open'. The first verb specifies a particular type of action (e.g., snapping, cutting with scissors, tearing), and often implies a state change (e.g., being snapped, cut, or torn), but does not assert that the state change actually came about; it is the second verb that confirms this, usually adding more information about the nature of the separation. For these languages we entered both verbs of the V-V compounds into the analysis.

Most problematic were classic "satellite-framed" languages like English, German, and Swedish (Talmy 1991), in which verbs are often accompanied by a prefix or a particle or other complement (*cut off*, *smash to bits*). According to Talmy, it is the satellite in such languages that characteristically encodes state change, while the main verb encodes only a "supporting event", such as a causal action or the manner of the state change. But all such languages in our sample in fact have a rich set of C&B verbs—e.g., English *cut*, *break*, *tear*, *smash*, and *snap*—whose meaning inherently entails the state change; the satellites mostly simply reinforce or further specify this state change. These verbs were clearly the event-encoding elements most comparable in meaning to the single verbs of languages like Yélî Dnye and to the first, action-specifying verbs in Mandarin V-V compounds. We decided, then, to focus on these verbs in our analysis of the categorization of C&B events, and to leave aside the satellites on grounds that they are semantically much more general than our target semantic domain.

3.2. *Analysis*

Our central question is how similar the semantic categories of C&B events are across languages. To determine this, we compared the extensions of the verbs elicited from speakers of the various languages of our sample, asking to what extent these verbs group and distinguish the C&B videoclips in the same way. According to our logic, the information available from linguistic event descriptions can be viewed as analogous to the data obtained in a psychologist's similarity judgment task. In such a task participants might be asked to sort stimuli—e.g., pictures on a set of cards—into groups of items they consider "similar". Any two cards are assigned a similarity score based on how often they end up in the same groups: if everyone puts them in the same group they are maximally similar, if no one does they are maximally dissimilar, and if some do and others do not they are intermediate in similarity. Similarity scores are calculated for all possible pairs of items, and this data structure, known as a (dis)similarity or proximity matrix, can be analyzed with multivariate statistical techniques (e.g., multidimensional scaling, correspondence analysis, clustering) to reveal the underlying dimensions or features according to which items are judged as similar or different.

Consultants in our study were not asked to sort or classify videotaped event clips, only to describe them. But each predicate they produced (*He cut a carrot; He broke a stick over his knee*) can be seen as defining an event category for them, equivalent to a card group. Clips that, across speakers, are often described with the same predicate (i.e., sorted into the same group) are more similar, from the semantic point of view, than clips typically described with different predicates. This approach to calculating similarity is useful for cross-linguistic data because it captures the extent to which speakers within and across languages describe any two clips with the same verb(s), regardless of the specific words they used.

Ideally, we would have many consultants from each language to allow a fine within-language calculation of intermediate degrees of similarity between event clips. Ideally too, we would have the same number of consultants for each language, since otherwise languages with more consultants could influence the representation of similarity among the clips more than languages with fewer consultants. But we had different numbers of consultants for different languages, with a few represented by only one or two speakers. To ensure that all the languages contributed equally to the analysis, we used an "all-or-none" scoring procedure that disregards intermediate degrees of similarity: two clips were scored as completely similar for a particular language if at least one consultant described them with the same verb, and as completely dissimilar if no consultant did (i.e., a binary value, 1 or 0, was assigned to the clip pair for that language, depending on whether the pair members were ever described with the same verb).

With this procedure a similarity matrix was constructed for each of the 28 languages separately (61 clips, taken pairwise, for a total of 1830 pairs in each matrix), and these were then jointly submitted to a correspondence analysis (Greenacre 1984) to extract the main dimensions that structure the semantic categorization of C&B events across languages (see Majid et al., forthcoming, for details of the statistical procedure). Correspondence analysis produces a "semantic map" that plots stimuli (here, the videoclips) in a multidimensional space, with the distance between any two stimuli reflecting their degree of similarity (here, the degree to which speakers of various languages used the same verb to describe them), calculated across the data set as a whole.

The dimensions are extracted in their order of importance, with the first dimension accounting for the most variance, the second for the next most, and so on. But the dimensions do not necessarily form an implicational

(AutoPDF V7 15/6/07 10:43) WDG (148×225mm) TimesM J-1787 COGL, 18:2 HC1: WSL(A) 14/06/2007 pp. 133–152 1787_18-2_01 (p. 141)

hierarchy; that is, a language can make a distinction along dimension three without necessarily also making one along dimensions two or one (cf. Majid et al., forthcoming). Following convention, we interpret only the first few dimensions.

3.3. *Results*

Our first analysis is based on data from all the videoclips. The first few dimensions extracted in this analysis neatly distinguished separation events such as opening, taking apart, and peeling from our set of putatively core C&B events that involve separation with material destruction (see Appendix). This means that events of opening, etc., are encoded the most distinctly across languages, in the sense that they rarely share verbs with C&B events or with each other (for interesting exceptions, see articles by van Staden and Gaby, this issue), whereas C&B events often share verbs with each other. Recall that we included events of opening and so on among our videoclips because—given children's errors—it was unclear a priori whether languages would agree on where C&B events leave off and other kinds of events begin. Apparently they do agree: across languages, C&B events hang together as a relatively coherent set, distinct from events involving other kinds of separation.

To get a better view of the inner structure of the C&B set, we performed a second correspondence analysis based only on these clips, omitting the clips showing opening, taking apart, and peeling.[2] A three-dimensional plot resulting from this analysis, with the clips plotted on the basis of their loadings on each of the dimensions, is shown in Figure 1.

Along the first dimension (shown on the x-axis), clips are strung out continuously (clip numbers have been slightly separated to make the plot easier to read). This dimension seems to capture the relative predictability of the locus of separation in the acted-on object. For clip 10 (far left: slicing carrots with a small knife), this locus is highly predictable: the separation will take place exactly where the knife is placed. For clip 40 (far right: smashing a plate with a hammer), predictability is poor: the plate may fracture in any place and fly into any number of pieces. For clip 48 (midway along the dimension: chopping off a branch with an axe), predictability is intermediate: the branch will separate lawfully where the axe blade falls, but since the axe swing is ballistic the locus can be predicted only within a margin of error.[3]

The extremes of dimension 1 are distinguished in all the languages; that is, the events represented to the far left are systematically described with different verbs than those represented to the far right. Clips at intermediate

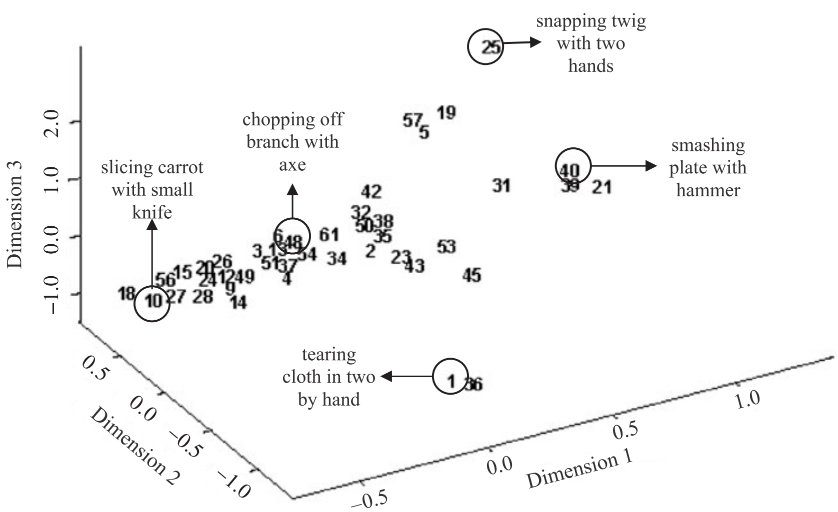


Figure 1. *A three dimensional plot of C&B events*

positions along the dimension are treated in different ways by different languages. Consider 48 (chopping off branch with axe) and 42 (breaking stick with karate chop). In English such events were often assigned to a category of their own (*chop*), but they were also often classed with events located toward the *left* end of the dimension (e.g., described with *cut*). The common denominator of this categorization is the use of a blade (-like) instrument to effect the separation, regardless of whether it is applied with precise placement or a ballistic swing. An opposing solution is adopted by German: chopping events were routinely classed with events positioned toward the *right* end of the dimension (described with *slagen*). The common denominator defining this category is the use of a sharp blow, whether by a bladed implement or a blunt one (Majid et al., this issue; see also Narasimhan, this issue; on fine-grained differences in the way Hindi and Tamil negotiate a similar boundary).

Another revealing comparison along dimension 1 involves English and the Caribbean Creole Sranan. As we have just seen, English weights instrument heavily along this continuum: separations brought about by a bladed implement (represented to the left, and often called *cut*) are distinguished from separations brought about in other ways (represented to the right, often called *break*). Sranan partitions dimension 1 with words derived from English *cut* and *break* (*koti* and *broko*), but the division between them falls further to the right. This is because Sranan cares more about the nature of the separation than about the

(AutoPDF V7 15/6/07 10:43) WDG (148×225mm) TimesM J-1787 COGL, 18:2 HC1: WSL(A) 14/06/2007 pp. 133–152 1787_18-2_01 (p. 143)

instrument: *koti* describes events involving a clean, cut-like fracture, whether effected by a blade or in some other way (Essegbey, this issue). English groups the intermediately-positioned clip 38 (*break*ing thread with hands—a clean fracture but no blade) with clip 39, to its right (*break*ing a pot with a hammer—no clean fracture, no blade), distinguishing them both from clip 10, to the left (*cut*ting carrot with knife—clean fracture, blade). Sranan, in contrast, groups 38 with 10 (both *koti*—clean fracture, blade or no blade) and distinguishes them from 39 (*broko*—messy fracture).

Dimension 2 (y-axis) distinguishes clips 1 (tearing cloth in two by hand) and 36 (tearing cloth halfway through) from all the other clips. Both involve a hand action on a two-dimensional flexible object. These "tearing" clips are located in the middle of the previous dimension, and indeed, tearing by hand, like chopping, seems intermediate in the predictability of the locus of separation. Many languages in the sample have a verb like English *tear*, which was used only for these two clips. But Yélî Dyne has a category that in our sample is unique: the verb used for the two tearing clips was also used for several carrot-cutting clips. According to Levinson (this issue), this verb picks out events of "severing with the grain" (with or without a bladed instrument).

Among events already distinguished by dimension 1 as having a poorly predictable locus of separation (to the far right of the x-axis), dimension 3 (z-axis) makes a further distinction between smashing a rigid object like a plate, pot, or carrot with a sharp blow (clips 21, 31, 39, 40) and snapping a long object like a stick or a carrot into two pieces between the hands or over the knee (clips 5, 19, 25, 57). Some languages respect this distinction perfectly (e.g., Ewe in Ameka and Essegbey, this issue), while others disregard it entirely (e.g., Hindi and Tamil use a general break verb for events of both kinds; Narasimhan, this issue). Still other languages offer a choice between a verb that groups events of the two types (e.g., English *break*, which applies to both) and verbs that distinguish them (e.g., English *snap* versus *smash*).

Do the three dimensions just discussed, and the positioning of the stimuli within the semantic space they define (as shown in Figure 1), capture the structure of individual languages in the sample, or does this solution arise simply as an artifact of averaging across languages? To explore this question we examined, for each dimension, how well the solution for individual languages correlates with the solution found when aggregating over all the languages (Majid et al., forthcoming). The mean correlations are high (.83, .79, and .62 for dimensions 1, 2, and 3), and the standard deviations are low. (It is logical that the correlation is highest for dimension 1 and lowest for 3, since, as discussed earlier, earlier-extracted dimensions explain more variance than later-extracted dimensions). This means that overall, the semantic space shown in Figure 1 does a good job of representing the structure of individual languages as well as of all the languages taken together.

4. Discussion

In this study we have seen that the extension of lexical categories used to describe C&B events across the languages of the world can be captured by a small number of dimensions. Events can be distinguished on the basis of how predictable the location of separation is in an object (dimension 1), tearing events are very often honored with a verb of their own (dimension 2), and snapping events are likely to be distinguished from smashing events (dimension 3).

Of course, none of the languages categorized C&B events in exactly the same way. For instance, there were enormous differences in the raw number of lexical categories into which speakers of different languages sorted the C&B clips, with Yélî Dyne speakers using only three verbs to describe the entire set (Levinson, this issue), and Tzeltal speakers using more than fifty (Brown, this issue; see also Palancar, this issue, on Otomi). There were also striking differences in the placement of category boundaries—cf. the partitioning of dimension 1 by English versus German, and by English versus Sranan, as discussed above. But this kind of variation plays out within a constrained space, and it can be well captured with a limited number of dimensions.

In closing, let us consider a lurking theoretical question: does our approach show something about the meaning of C&B verbs? Admittedly we have discussed only the *extensions* of verbs, and have not tried to abstract away to more formal *intensional* representations of the sort linguists and psychologists often attempt—e.g., feature-based compositional representations such as 'x CAUSE [y BECOME **BROKEN**]' (for transitive *break* in English), and 'x produce **CUT** in y, by sharp edge coming into contact with y' (for *cut*) (Guerssel et al. 1985; Jackendoff 1990; Levin and Rappaport 1995; Pinker 1989). In the semantic domain of C&B as with other kinds of caused state change, a typical compositional representation involves two basic types of features: *primitive predicates* such as CAUSE and BECOME and so-called *constants* such as BROKEN (Levin and Rappaport 1995: 23). Constants carry a tremendous burden since they are all that distinguishes the meaning of e.g., *break* from the meaning of e.g., *smash* or *snap*. Yet constants are essentially black boxes, since so far, there has been no account of what it actually *means* to break something or to smash something. For example, how can constants capture

(AutoPDF V7 15/6/07 10:43) WDG (148×225mm) TimesM J-1787 COGL, 18:2 HC1: WSL(A) 14/06/2007 pp. 133–152 1787_18-2_01 (p. 145)

the systematicity behind the fact that English *break* covers thread-breaking but Sranan *broko* does not? Or that English *break* covers both stick-snapping and pot-smashing while its Dutch cognate *breken* is limited to snapping?[4]

For contemporary usage-based approaches to language, it is essential to be able to pinpoint the boundaries of the extensions of cognate or similar words in different languages, and to document their slow diachronic creep as they gain or lose territory from/to their competitors over time. It is precisely in its ability to capture such information, and so to reveal what remains obscured by constants, that the extensional approach shows its power.

Received 17 February 2006
Revision received 8 December 2006

Max Planck Institute for Psycholinguistics, The Netherlands
University of Amsterdam, The Netherlands
University of Connecticut, USA

Appendix

Below are short descriptions of the video stimuli used in this study, designed by Bohnemeyer, Bowerman, and Brown (2001). The videoclips are distinguished according to whether they are agentive. Clips showing an agent appear in normal font, spontaneous events with no agents are shown in bold-face. Italics indicate the open, take apart, and peel items that were extracted by the first and second factors of the initial correspondence analysis. Bold and italic items are omitted from the analyses presented in Figure 1.

1. Tear cloth into two pieces by hand
2. Cut rope stretched between two tables with single downward blow of chisel
3. Hack branch off tree with machete
4. Chop cloth stretched between two tables with repeated intense knife blows
5. Break stick over knee several times with intensity
6. Chop multiple carrots crossways with big knife, intensity
7. *Push chair back from table*
8. **Piece of cloth tears spontaneously into two pieces**
9. Slice carrot lengthwise with knife into two pieces
10. Slice carrot across into multiple pieces with knife

11. Pull two paper cups apart by hand
12. Cut strip of cloth stretched between two people's hands in two
13. Cut rope stretched between two tables with blow of axe
14. Make single incision in melon with knife
15. Saw stick propped between two tables in half
16. Forking branch of twig snaps spontaneously off
17. Carrot snaps spontaneously
18. Cut finger accidentally while cutting orange
19. Snap twig with two hands
20. Cut single branch off twig with sawing motion of knife
21. Smash carrot into several fragments with hammer
22. Take top off pen
23. Chop cloth stretched between two tables into two pieces with two blows of hammer
24. Cut rope in two with scissors
25. Snap twig with two hands, but it doesn't come apart
26. Cut carrot crossways into two pieces with a couple of sawing motions with knife
27. Cut hair with scissors
28. Cut fish into three pieces with sawing motion of knife
29. Peel an orange almost completely by hand
30. Peel a banana completely by hand
31. Smash a stick into several fragments with single blow of hammer
32. Cut carrot in half crossways with single karate chop of hand
33. Open a book
34. Chop cloth stretched between two tables with single karate chop of hand
35. Break yarn into many pieces with fury
36. Tear cloth about half-way through with two hands
37. Cut carrot in half lengthwise with single blow of axe
38. Break single piece off yarn by hand
39. Smash flower pot with single blow of hammer
40. Smash plate with single blow of hammer
41. Open a hinged box
42. Break vertically-held stick with single karate chop of hand
43. Cut carrot crossways into two pieces with single blow of chisel
44. Open cannister by twisting top slightly and lifting it off
45. Poke hole in cloth stretched between two tables with a twig
46. Rope parts spontaneously, sound of a single chop
47. Open hand
48. Chop branch repeatedly with axe, both lengthwise and crosswise, until a piece comes off

(AutoPDF V7 15/6/07 10:43) WDG (148×225mm) TimesM J-1787 COGL, 18:2 HC1: WSL(A) 14/06/2007 pp. 133–152 1787_18-2_01 (p. 147)

49. Cut rope in two with knife
50. Chop rope stretched between two tables in two with repeated blows of hammer
51. Split melon in two with single knife blow, followed by pushing halves apart by hand
52. Open mouth
53. Break stick in two with single downward blow of chisel
54. Cut carrot in half crosswise with single blow of axe
55. Open teapot/take lid off teapot
56. Cut cloth stretched between two tables in two with scissors
57. Snap carrot with two hands
58. Open eyes
59. Open scissors
60. Open door
61. Break rope stretched between two tables with single karate chop of hand

Notes

* We thank Stephen Levinson, Nick Enfield, Loretta O'Connor and all the members of the Language and Cognition group for their input to this introduction and to the formulation of the CUT AND BREAK project reported in this special issue. The papers in this volume benefited from critical and thoughtful feedback from the reviewers, who we would also like to acknowledge. Finally, we are very grateful to Marloes Huijbers for her expert and patient assistance in preparing the volume, and Ludy Cilissen for his work in preparing many of the illustrations. Any correspondence should be addressed to Asifa Majid, Max Planck Institute for Psycholinguistics, Postbus 310, Nijmegen, 6525XD, The Netherlands, or email to ⟨Asifa.Majid@mpi.nl⟩.

1. The video clips are available on request from ⟨Asifa.Majid@mpi.nl⟩.
2. We also omitted the four clips showing spontaneous (non-agentive) C&B events (see Appendix); these had been included in the elicitation tool to explore questions of argument structure (see Bohnemeyer, this issue) and are not relevant for present purposes.
3. All the clips in this analysis showed an agent, so predictability corresponds closely to the agent's degree of control over the locus of separation (Majid et al. 2004). But we emphasize predictability rather than control because the verbs that are associated with a particular region of dimension 1—e.g., English *cut*, to the left—are applicable even if an agent acts unintentionally (e.g., inflicts an accidental cut, clip 18).
4. Goddard (1998) offers an alternative decompositional approach using "natural semantic metalanguage", a system that represents the meanings of words with a set of semantic primes. Goddard provides a (partial) analysis of transitive *break* as: 'x did something, because of this, something happened to y at this time, because of this, after this y was not one thing anymore'. This approach suffers from some of the same problems as the classical decompositional approach since it does not, as yet, offer a fine enough resolution to distinguish between the various ways in which "not being one thing anymore" can be interpreted, or in which this state of affairs can come about.

References

Ameka, Felix and James Essegbey
this issue CUT and BREAK verbs in Ewe and the causative alternation construction. *Cognitive Linguistics* 18(2), 241–250.

Bloom, Paul
2000 *How Children Learn the Meaning of Words*. Cambridge, MA: MIT Press.

Bohnemeyer, Jürgen, Melissa Bowerman and Penelope Brown
2001 Cut and break clips. In Levinson, Stephen C., and N. J. Enfield (eds.), *Field Manual 2001, Language and Cognition Group, Max Planck Institute for Psycholinguistics.* Nijmegen: MPI, 90–96.

Bohnemeyer, Jürgen
this issue Morpholexical relatedness and the argument structure of verbs of CUTTING and BREAKING. *Cognitive Linguistics* 18(2), 153–177.

Bowerman, Melissa
1989 Learning a semantic system: What role do cognitive predispositions play? In Rice, Mable L. and Richard L. Schiefelbusch (eds.), *The Teachability of Language*. Baltimore: Brooks, 133–169.
1996 The origins of children's spatial semantic categories: Cognitive vs. linguistic determinants. In Gumperz, John J. and Stephen C. Levinson (eds.), *Rethinking Linguistic Relativity*. Cambridge: Cambridge University Press, 145–176.
2005 Why can't you 'open' a nut or 'break' a cooked noodle? Learning covert object categories in action word meanings. In Gershkoff-Stowe, Lisa, and David Rakison (eds.), *Building Object Categories in Developmental Time: 32nd Carnegie Symposium on Cognition*. Mahwah, NJ: Lawrence Erlbaum, 33–62.

Bowerman, Melissa and Soonja Choi
2001 Shaping meanings for language: universal and language-specific in the acquisition of spatial semantic categories. In Bowerman, Melissa and Stephen C. Levinson (eds.), *Language Acquisition and Conceptual Development*. Cambridge: Cambridge University Press, 475–511.
2003 Space under construction: Language-specific spatial categorization in first language acquisition. In Gentner, Dedre and Susan Goldin-Meadow (eds.), *Language in Mind*. Cambridge, MA: MIT Press, 387–427.

Brown, Penelope
this issue 'She had just cut/broken off her head': CUTTING and BREAKING verbs in Tzeltal. *Cognitive Linguistics* 18(2), 319–330.

Chen, Jidong
this issue "He *cut-break* the rope": Encoding and categorizing CUTTING and BREAKING events in Mandarin. *Cognitive Linguistics* 18(2), 273–285.

Clark, Eve V.
1976 Universal categories: On the semantics of classifiers and children's early word meanings. In Juilland, A. (ed.), *Linguistic Studies Offered to Joseph Greenberg on the Occasion of his Sixtieth Birthday*. Saratoga, CA: Anna Libri, 449–462.

Clark, Herb H.
1973 Space, time, semantics, and the child. In Moore, T. E. (ed.), *Cognitive Development and the Acquisition of Language*. New York, NY: Academic Press, 27–64.

de León, Lourdes
2001 Finding the richest path: Language and cognition in the acquisition of verticality in Tzotzil (Mayan). In Bowerman, Melissa and Stephen C. Levinson (eds.), *Language Acquisition and Conceptual Development*. Cambridge: Cambridge University Press, 544–565.

Dryer, Matthew S.
1989 Large linguistic areas and language sampling. *Studies in Language* 13, 257–92.

Essegbey, James
this issue CUT and BREAK verbs in Sranan. *Cognitive Linguistics* 18(2), 231–239.

Gaby, Alice
this issue Describing CUTTING and BREAKING events in Kuuk Thaayorre. *Cognitive Linguistics* 18(2), 263–272.

Gleitman, Lila
1990 The structural sources of verb meanings. *Language Acquisition* 1, 3–55.

Goddard, Cliff
1998 *Semantic Analysis: A Practical Introduction*. Oxford: Oxford University Press.

Greenacre, Michael J.
1984 *Theory and Applications of Correspondence Analysis*. London: Academic Press.

Guerssel, Mohamed, Kenneth Hale, Mary Laughren, Beth Levin, and Josie White Eagle.
1985 A cross-linguistic study of transitivity alternations. In Eilfort, William H., Paul D. Kroeber, and Karen L. Peterson (eds.), *Papers from the Parasession on Causatives and Agentivity at the 21st Regional Meeting*. Chicago: Chicago Linguistic Society, 48–63.

Hale, Kenneth L. and Samuel J. Keyser
1987 *A View from the Middle*. (Lexicon Project Working Papers 10.) Cambridge, MA: Center for Cognitive Science, MIT Press.

Jackendoff, Ray
1990 *Semantic Structures*. Cambridge, MA: MIT Press.

Levin, Beth
1993 *English Verb Classes and Alternations.* Chicago: University of Chicago Press.

Levin, Beth, and Malka Rappaport Hovav
1995 *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Cambridge, MA: MIT Press.

Levinson, Stephen C.
this issue CUT and BREAK verbs in Yélî Dnye, the Papuan language of Rossel Island. *Cognitive Linguistics* 18(2), 207–217.

Majid, Asifa, James S. Boster and Melissa Bowerman
Forthc. The cross-linguistic categorization of everyday events: A study of CUTTING and BREAKING.

Majid, Asifa, Miriam van Staden, James S. Boster, and Melissa Bowerman
2004 Event categorization: a cross-linguistic perspective. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*. Chicago, IL: Cognitive Science Society, 885–890.

Majid, Asifa, Marianne Gullberg, Miriam van Staden and Melissa Bowerman
this issue How similar are semantic categories in closely related languages? A comparison of four Germanic languages. *Cognitive Linguistics* 18(2), 179–194.

Malt, Barbara, Steven Sloman and Silvia Gennari
2003 Universality and language specificity in object naming. *Journal of Memory and Language* 49, 20–42.

Malt, Barbara, Steven Sloman, Silvia Gennari, Meiyi Shi and Yuan Wang
1999 Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language* 40, 230–262.

Narasimhan, Bhuvana
this issue CUTTING, BREAKING, and TEARING verbs in Hindi and Tamil. *Cognitive Linguistics* 18(2), 195–205.

Palancar, Enrique
this issue CUTTING and BREAKING verbs in Otomi: An example of lexical specification. *Cognitive Linguistics* 18(2), 307–317.

Perkins, R. D.
1989 Statistical techniques for determining language sample size. *Studies in Language* 13, 293–315.

Piaget, Jean
1954 *The Construction of Reality in the Child*. New York, NJ: Basic Books.

Pinker, Steven
1989 *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: MIT Press.

Pye, Clifton
1994 Breaking concepts: Constraining predicate argument structure. Unpublished manuscript. Department of Linguistics, University of Kansas.
1996 K'iche' Maya verbs of breaking and cutting. In Goodell, Melissa and Dong Ik Choi (eds.), *Kansas Working Papers in Linguistics 21*. Lawrence, KS: University of Kansas, 87–98.

Pye, Clifton, Diane Frome Loeb and Yin-Yin Pao
1995 The acquisition of BREAKING and CUTTING. Paper presented at The Twenty-Seventh Annual Child Language Research Forum.

Rijkhoff, J. and D. Bakker
1998 Language sampling. *Linguistic Typology* 2, 262–314.

Ross, Malcolm
1996 Contact-induced change and the comparative method: Cases from Papua New Guinea. In Durie, Mark and Malcolm Ross (eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, Oxford: Oxford University Press, 180–217.

Schaefer, Ronald P.
1979 Child and adult verb categories. In *Kansas Working Papers in Linguistics* 4, Lawrence, KS: University of Kansas, 61–76.

Slobin, Dan I.
1973 Cognitive prerequisites for the development of grammar. In Ferguson, C. A. and D. I. Slobin (eds.), *Studies of Child Development*. New York: Holt, Reinhart and Winston, 175–208.

Talmy, Leonard
1985 Lexicalization patterns: Semantic structure in lexical form. In Shopen, Timothy (ed.), *Language Typology and Syntactic Description*, vol. III*: Grammatical Categories and the Lexicon*. Cambridge, UK: Cambridge University Press, 57–149.
1991 Path to realization: A typology of event conflation. In *Proceedings of the Berkeley Linguistics Society* 17, 480–519.

(AutoPDF V7 15/6/07 10:43) WDG (148×225mm) TimesM J-1787 COGL, 18:2 HC1: WSL(A) 14/06/2007 pp. 133–152 1787_18-2_01 (p. 151)

Toth, N. and K. Schick
1993 Early stone industries and inferences regarding language and cognition. In Gibson, K. R. and T. Ingold (eds.), *Tools, Language and Cognition in Human Evolution*. Cambridge: Cambridge University Press, 346–362.
van Staden, Miriam
this issue Please open the fish: Verbs of separation in Tidore, a Papuan language of Eastern Indonesia. *Cognitive Linguistics* 18(2), 297–306.
Wilkins, David P. and Deborah Hill
1995 When "go" means "come": Questioning the basicness of basic motion verbs. *Cognitive Linguistics* 6, 209–259.