



PART ONE

General Overview of Elicitation Techniques





1

A Practical Epistemology for Semantic Elicitation in the Field and Elsewhere

Jürgen Bohnemeyer

1 Introduction

The aim of this chapter is to sketch a classification of elicitation methods in semantics based on an analysis of the sources of evidence semanticists can draw on and the principal components of any elicitation.*

As part of a general empiricist turn that has been slowly but inexorably changing the cognitive sciences for the last couple of decades, linguists working in the classical core areas of morphosyntax, semantics, and phonology have begun to feel the need for a more empirical footing of their research. The current concern with variation and the growing importance of statistical methods throughout these subfields seem to be testimony of this.

Furthermore, I believe that the tools for a more empirically grounded linguistics are by and large already in our hands. Linguists have developed a host of procedures for data gathering and analysis since the days of the Structuralists. What is called for now is an integration of these methodologies with the epistemological standards of the social and behavioral sciences such as those that sociolinguists and psycholinguists have been adhering to in their research. This chapter attempts to make a modest contribution toward this general goal in the area of semantic research.

The temptation of thinking of semantics as necessarily hermeneutically based is perhaps greater than in other subfields of linguistics. After all, meaning is not

*Different parts of the discussion in this chapter are based upon work supported by the National Science Foundation under Grant No. BCS-0723694 “Spatial language and cognition in Mesoamerica” (PI J. Bohnemeyer) and by the Max Planck Society, respectively. I owe a wealth of gratitude to the Yucatec speakers who participated in the studies discussed here. I would like to thank Rui Chaves, Florian Jaeger, Jean-Pierre Koenig, and Gail Mauner for helpful suggestions. I am indebted to the editors and Judith Tonhauser for detailed, engaging, and constructive comments. I alone am to blame for the positions argued for and for any shortcomings.

14 Methodologies in Semantic Fieldwork

directly observable—so how else would it be accessible if not through interpretation? This fallacy is directly responsible for the belief Matthewson (2004) dubbed “relativist agnosticism”: the widespread assumption that it is impossible to study the semantics of languages the researcher does not speak, or at least does not speak native-like. If meaning is accessible through interpretation only, then how could one study semantics in a language the utterances of which one does not know how to adequately interpret? For this reason, research methods for semantics in field contexts are of primary concern in this chapter—which is, of course, why this chapter is a part of the present volume. But the same apparent conundrum presents itself in child language research: to the extent that the linguistic competence and practice of children is not adult-like, how could adult researchers understand children’s utterances well enough to adequately interpret their semantics?

This apparent conundrum is nothing more than the everyday reality of every psychologist: mental states and processes are not directly observable, they must be inferred from the behaviors they are assumed to underlie, in particular, participants’ responses to particular stimuli in controlled research settings. I contend that this is a fairly close analogy to how we should think of semantic research. It is from this perspective that the survey of semantic elicitation techniques in this chapter is offered.

The primary guiding maxim is that an empirical linguist cannot rely on their own native speaker intuitions as their sole source of evidence. This principle places the semantic field researcher in the same epistemological boat as any other empirically oriented semanticist. Like the child learning the meanings of the expressions of her native language, as modeled in Brown’s (1958) “Original Word Game” (see Figure 1.1), they must start out observing utterances produced by competent speakers and their apparent referential correlates in the observable world. This holds true even for semanticists who view meaning primarily as a relation between utterances and the mind rather than between utterances and the world, as mental states are not directly observable. The utterance–world correlations can be augmented by utterance–utterance correlations, including metalinguistic utterances—for example, judgment statements—and contact language utterances. These must not be naively taken to simply represent the meanings of the target language utterances; they can, however, serve as further data points in a correlation matrix.

Again like the child in the Original Word Game, the semanticist must now formulate, based on the observed correlations, hypotheses concerning the meanings of the utterances and their constituents and then proceed to test these. The central questions of an empirical epistemology and methodology for semantics are the following:

- (1) Which observable properties of communicative behavior can be exploited as the data of semantic research?
- (2) How and under what constraints can evidence of these observable properties be gathered in a fashion that permits valid analyses?

- (3) Which analytical techniques are appropriate and optimal for bringing a particular dataset to bear on a particular research question?

This chapter is primarily concerned with the second question. But since any meaningful answer to (2) presupposes the existence of plausible and practicable answers to (1), I briefly consider it in section 3.

Against the backdrop of the general picture of semantic research rooted in the observation of communicative behavior, the present chapter develops a classification of data gathering techniques that can serve as possible solutions in response to question (2). The focus will be on methods for semantic elicitation. Elicitation can be defined as a data collection technique that involves three principal components: a stimulus, a task, and a response. In any kind of linguistic elicitation, the stimulus may be a target language utterance; a contact language utterance; a linguistic representation of some state of affairs (e.g., a description of some scenario the native speaker consultant is asked to assume); a non-linguistic representation of some state of affairs; or a combination of any of the above. The response may consist of a target language utterance produced by the speaker; a judgment that may form the basis for diagnoses of well-formedness, truth conditions, and so forth; or again a linguistic (e.g., explication by paraphrase) or a non-linguistic (e.g., in demonstrations and act-out tasks) representation of some state of affairs. All possible tasks may then be defined as mappings between possible stimuli and possible responses. I argue that there are only seven possible types of elicitation techniques in linguistics. Applications of all of these to semantic research will be illustrated with examples from my fieldwork.

As the semanticist must observe native speaker intuitions about utterance-world correlations as their primary source of evidence, linguistic and non-linguistic representations as stimuli play a powerful role in semantic elicitation since they allow the researcher to manipulate what the world is assumed to be like for the study of particular utterances. Finally, I propose a Golden Rule of elicitation: the validity of any elicitation response as a data point in the reconstruction of the speaker's linguistic competence depends on the speaker's interpretation of the task and the stimulus and their intended interpretation of their response.

The chapter is organized from here on as follows: I first elaborate on the answer to question (1) I offered above. I then introduce the proposed classification of data gathering techniques and specifically that of elicitation techniques and proceed to illustrate the various types with examples from my field research on Yucatec Maya. A discussion section introduces the Golden Rule and revisits the dichotomy between empirical and hermeneutic research.

2 The Empirical Basis of (Field) Semantics

Empirical research in all areas of linguistics has been making strides, including in semantics. Nevertheless, an explicit epistemological and methodological footing

for an empirical approach to the study of meaning has been lacking. Laypeople and many trained linguists alike continue to assume that interpretation is the only route to meaning and that semantics is therefore necessarily a hermeneutic enterprise. For the study of meaning in contexts where hermeneutic approaches alone are insufficient—for example, in field research on indigenous languages, in work with aphasic patients, and in the study of meaning in child language and gesture—this belief system is presenting a serious obstacle. It is obviously not a belief system held by the contributors to this volume—but many other colleagues, both fieldworkers and non-fieldworkers, do seem to subscribe to it.

A principal challenge for an empirical science of meaning is the fact that the meanings of linguistic utterances are not directly observable. Instead, they must be inferred from the observable semiotic behavior of language users, which is to say from their communicative behavior. In this respect, the task of the empirical semanticist resembles that of the psychologist, who infers the properties of cognitive representations and processes from the observable behavior they are assumed to support. The semanticist's position is also not unlike that of a child acquiring semantic and pragmatic competence in the language she is exposed to. Figure 1.1 presents a cartoon version of the processes presumably involved in semantic acquisition and empirical research in semantics alike: the child/researcher

- ❑ Observes apparent correlations between utterances (A) and stimuli (C);
- ❑ Derives inferences concerning the meaning of the utterance and its constituents (B);
- ❑ Formulates hypothesis about the semantic system in the mind of the observed competent speakers (the arrow in the bottom right corner of Figure 1.1);
- ❑ Tests these hypotheses by manipulating the stimulus (or waiting for a variation of it to occur spontaneously) and observing how this affects the response (the arrow in the bottom left corner of Figure 1.1).

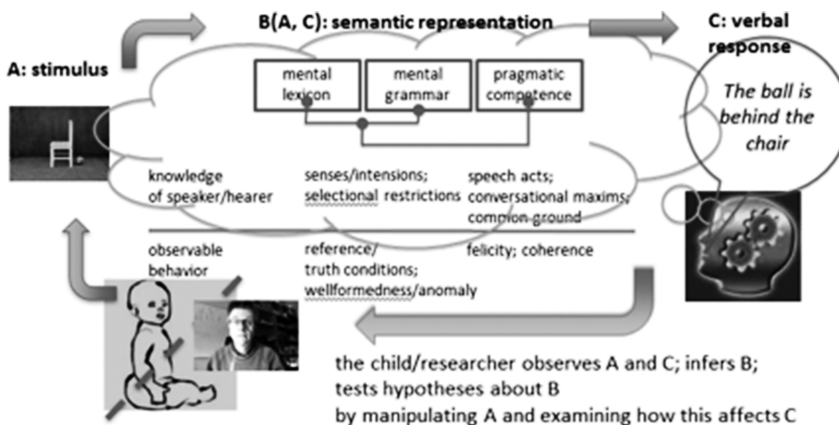


FIGURE 1.1 An elaborate update of Brown's (1958) "Original Word Game"

The richest and most basic source of observational evidence in semantics is the extensions of linguistic signs: their sets of possible referents. While extensions cannot be observed exhaustively, since they are not finite sets, language learners and semanticists can observe individual entities or states of affairs that speakers use the words to label, can form hypotheses about the intensions or senses governing membership in the extensions, and can test these hypotheses, for example, by attempting to label new items and observing the response of competent speakers. The reference of such expressions can also be studied by observing their impact on the reference of the phrases and sentences they occur in. A special case of this is the contribution the various constituents of declarative sentences that serve to assert propositions make to the truth conditions of such sentences. For a simple illustration of how the connection between reference and truth conditions can be exploited in semantic research, here is a recipe for the study of the meaning of the preposition *under*: Take the sentence template in (4):

(4) *The x is under the y.*

Replace *x* and *y* with terms describing various objects. Show speakers spatial configurations of instances of the objects in question and ask them for each test configuration whether the sentence is true or false. Try to describe what all those configurations that make the sentence true according to the participants have in common.

Two concepts closely related to that of truth conditions are those of entailments and contradictions. Entailments and contradictions can be fairly straightforwardly correlated with observable communicative behaviors. For example, a sequence of two utterances that are incompatible in that one contradicts an entailment of the other will not only make it impossible for a native speaker consultant to come up with a coherent scenario that instantiates the sequence of utterances, but is likely to result more immediately in confusion and rejection on the part of the speaker. Thus, entailments and contradictions open up further alleys for empirically probing the truth conditions of given utterances. At the same time, predicting the truth conditions and entailments of particular utterances on the basis of their form and the meanings of their constituents is of course a central goal of semantic research.

Referential data is extensional. To get at the intensions or sense meanings—the conceptual content—of the elicited expressions, semantic and pragmatic analyses must be performed to separate entailments of lexical and compositional semantics from pragmatically generated meaning components. For a simple illustration, consider the responses to the Topological Relations Picture Series (affectionately known as “BowPed” after the authors, Melissa Bowerman and Eric Pederson (Bowerman and Pederson ms.)) by two Mexican Spanish speakers, represented by the Venn diagrams in Figure 1.2:

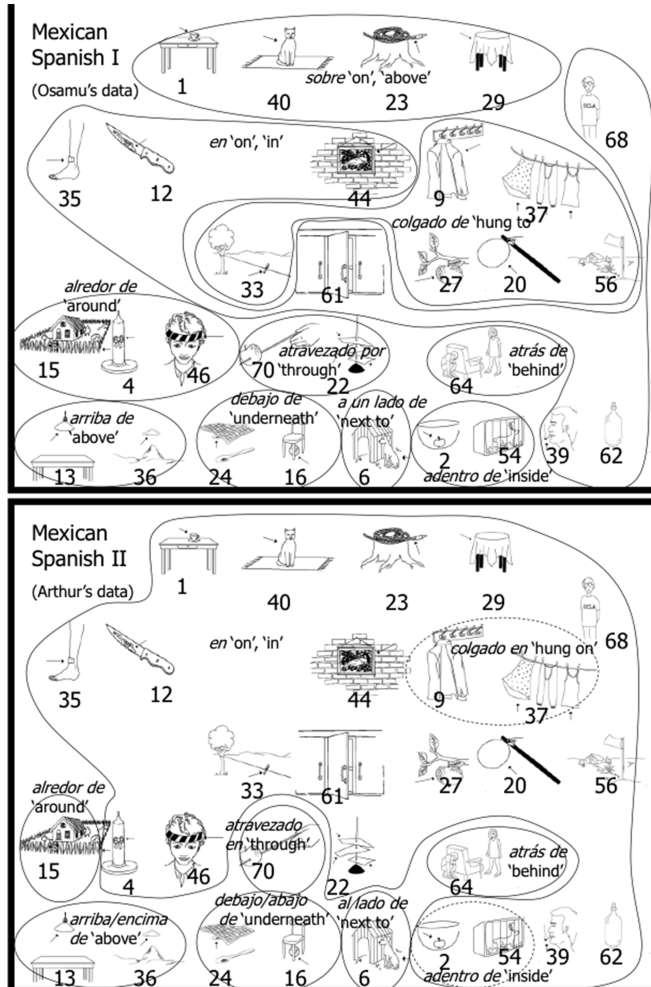


FIGURE 1.2 Linguistic categorization of a subset of the BowPed scenes by two Spanish speakers (data elicited by Osamu Ishiyama and Arthur Photidiadis)

It seems possible, but not likely, that these two speakers have different mental lexicon entries for the preposition *en*. A more parsimonious explanation is that the first speaker is restricting *en* to scenes to which more specific alternatives do not apply—a preemption effect based on a scalar implicature licensed by Grice’s first Maxim of Quantity, “Make your contribution as informative as is required.” This is an illustration of how conversational implicatures can make the extension of an expression appear narrower than it semantically is. Conversely, semantic transfer—metaphor and metonymy—may widen the expression’s extension.

Another indirectly observable property of the semantic extension of expressions is the potential variation in terms of prototypical and more marginal referents in the sense of Prototype Theory (Rosch and Mervis 1975; Rosch 1978). Many

expressions have structured extensions that include both focal and more marginal instances. This kind of gradation can be the result of vagueness (or underspecification), but also of membership in the extension being determined by similarity to a prototype or by the degree of satisfaction of a set of violable criteria (known as “preference rule systems” (Jackendoff 1983) or “idealized cognitive models” (Lakoff 1987)). Speakers are able to judge the goodness of a given referent and moreover often flag marginal instances in their spontaneous productions, using hedges—(e.g., when saying that a penguin is “*sort of a bird*”) (Lakoff 1973:471). Similarly, speakers are able to detect referents licensed by semantic transfer—metaphor or metonymy. However, this ability diminishes gradually as transferred uses become conventionalized. Thus, whereas a statement along the lines of (5) is likely to be considered true without much reflection, (6) is more likely to trigger a double take, because “hot dog,” unlike “toy dog,” is a fully conventionalized sense of the English word *dog*:

- (5) *A toy dog isn't actually/really a dog.*
 (6) *A hot dog isn't actually/really a dog.*

Furthermore, speakers have the ability to detect whether particular utterances are appropriate for their contexts. Inappropriateness can have a variety of sources, including anomaly, but also the failure to meet cultural norms of interaction, such as norms of politeness or cultural requirements of particular speech acts. Where these conditions are not met, the utterance is infelicitous and therefore fails to have the conventional effect it has in the right circumstances. Another kind of semantic condition that contexts must meet in order for certain utterances to be appropriately interpretable in them is presuppositions. Lastly, language-processing evidence can be tapped into in exploring the structure of the mental lexicon. For instance, the degree of semantic relatedness of two lexical items can be assessed by using word association tasks (Clark 1970). It is also accessible to priming effects in lexical decision (“Is the stimulus a word?”) or semantic categorization tasks (e.g., “Does the stimulus label a kind of animal?”); cf. Meyer, Schvaneveldt, and Ruddy (1972, 1974); Marslen-Wilson et al. (1994); Perea and Rosa (2002a, b); Ferrand and New (2003); Bueno and Frenck-Mestre (2008); inter alia. Interference effects in chronometric sentence processing (Lewis et al. 2006 and the literature cited there) and picture-naming tasks (Costa et al. 2005 and the literature cited therein) are likewise known to be sensitive to semantic relations. Even neurological evidence can in principal be brought to bear on the semantic relatedness of two expressions, including evidence from the impairment complex known as “deep dyslexia” (Marshall and Newcombe 1973; Plaut and Shallice 1993 and references therein). For further discussion on the use of psycholinguistic evidence in semantic studies, see Krifka 2011.

Important objections have been advanced against semantic externalism, referencing the potential problem of referential indeterminacy (Quine 1960) or citing a philosophical rejection of objectivism (Lakoff 1987; Jackendoff 2002). I take these criticisms extremely seriously. In my personal view of the relation between science and reality, I favor Constructive Empiricism (van Fraassen 1980) over Scientific

Realism. That is, I consider the measure of good science to be empirical adequacy, not objective truth. However, it seems to me that the task of inferring the semantic system of a language from the observation of the communicative behavior of its speakers must have a solution that is humanly attainable and that has an outcome that is sufficiently shared among the members of the speech community to allow the replication of the vast majority of this semantic system in the generations to which it is transferred. And if children can infer the semantics of a language from observing the behavior of competent speakers, I see no principal reason why semanticists should be unable to do the same, however different the tasks of the child and the semanticist are in every other respect.

3 Understanding Elicitation

Linguistic data gathering (in the broad sense, as opposed to just elicitation) involves maximally three components: a stimulus, a task, and a response. The stimulus is a linguistic or non-linguistic representation intended as the input of the task (cf. also Burton and Matthewson, this volume; Bochnak and Bogal-Allbritten, this volume; Bar-el, this volume). In comprehension and judgment tasks, the input (stimulus) is an utterance; in production tasks, the input (stimulus) constrains the content of the utterance to be produced. The semantic elicitation task (as opposed to the stimulus) is a speech act directed at the participant(s) by the researcher intended to trigger a set of computations involving the semantic system. These computations are intended to ultimately result in a response from which the computations, the representations, and the speaker's (or the speakers') knowledge and practices involved in them can then be recovered. In this way, the response permits inductive generalizations and the testing of hypotheses about the semantic system. The response is a communicative action in the broadest sense. It may be a target language utterance, a contact language translation, a metalinguistic judgment, or any non-linguistic action that solves the task, for example by pointing out a possible referent, demonstrating an action that would instantiate a given description, and so forth. Figure 1.3 presents a cartoon version of a rather pedestrian example: the field researcher asks a speaker of Yucatec Maya how to say "I've got

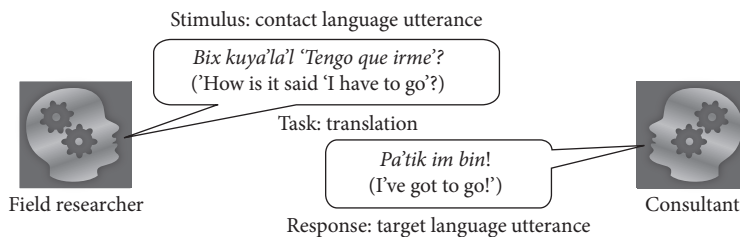


FIGURE 1.3 Components of linguistic data gathering

to go” in their native language, formulating the question—i.e., the task—in Maya, but the stimulus utterance in the contact language, Spanish. The speaker responds with the idiomatic Yucatec way of saying “I’ve got to go” (as an informal way of taking leave).

Not all of the three components are necessarily present in every study. There is an implicational relationship here: studies that employ stimuli require tasks, and all empirical studies of linguistic behavior examine acts of linguistic behavior—most commonly, utterances—whether these are responses to tasks and stimuli or not. We thus arrive at the classification in Table 1.1.

The plus and minus signs represent the presence and absence of the particular component, respectively. For the distinction between spontaneous and “staged” speech events, see Himmelmann (1998). I assume this distinction to be continuous. No speech event recorded by an observer is 100% spontaneous or staged. The greater the influence of the researcher, the more staged the event. Recordings of folk tales and descriptions of cultural practices—arguably the mainstay of linguistic field work in the Boasian tradition—typically exemplify the staged type, as the speakers realize the event in response to a request by the researcher (see also Cover and Tonhauser, this volume).

Linguistic elicitation then can be defined as the collection of responses to linguistic or non-linguistic stimuli designed to study the respondents’ linguistic competence and/or their practices of language use.¹ This yields a very broad notion of elicitation, going well beyond the traditional prototype of one speaker answering a researcher’s questions and including many techniques that are widely considered “experimental” rather than instances of elicitation. In my view, however, elicitation and experimentation are not mutually exclusive. Elicitation is an approach to data gathering. As such, it contrasts with recordings of spontaneous and staged speech events. Experimentation, on the other hand, is broadly any empirical test of a hypothesis and in the narrow sense involves observations under controlled

TABLE 1.1

The families of data gathering techniques in linguistics

	Recording of spontaneous speech events	Recording of staged speech events	Elicitation
Linguistic behavior (“response”)	+	+	+
Task	–	+	+
Stimulus	–	–	+

¹Experimental methods of psycholinguistics are by this definition not elicitation techniques, since the data they produce serve to test hypotheses about language processing, rather than to provide direct evidence of the participants’ linguistic competence and practices of language use. For this reason, psycholinguistic paradigms are not included in the further discussion. Nevertheless, as mentioned above, psycholinguistic data can be exploited for indirect clues about semantic relations.

22 Methodologies in Semantic Fieldwork

conditions. From this perspective, psycholinguistic experiments involve the elicitation of communicative behavior, and elicitation in turn may but need not be a part of an experiment, depending on whether it is conducted as a test of some hypothesis or merely for exploratory purposes.

A classification of the elicitation techniques at the linguist’s disposal can be achieved by crosstabulating the possible stimulus and response types and identifying the task types as mappings from the stimulus types into the response types, as depicted in Table 1.2. This presupposes that the identification of stimulus (“input”) and response (“output”) types alone is sufficient to define the task types. I can see no obvious theoretical reason why this should be so, but it seems to work out quite nicely.

The classification in Table 1.2 distinguishes four stimulus types. The stimulus is either an utterance—in which case it may be a target or a contact language utterance—or the content of some linguistic or non-linguistic representation. Of course, a stimulus utterance likewise conveys a particular meaning. The difference between an utterance used as stimulus and the content of a linguistic representation used as stimulus is that in the former case, the morphosyntactic and phonological properties, the particular set of lexical items involved, and the register are all part of the stimulus. The speakers’ response will be observed and analyzed as a response to all of these properties. In contrast, in the case of the content serving as stimulus, everything besides the meaning of the utterance is considered just “wrapping” and assumed as irrelevant to the speakers’ response (an assumption that may of course not always be borne out). An example of the content of

TABLE 1.2

A classification of elicitation techniques by stimulus and response type

Response stimulus	Target language utterance	Contact language utterance	Metalinguistic utterance: judgment	Metalinguistic utterance: description	Non-linguistic representation
Target language utterance	Type I—completion; association	Type II—translation	Type V—judgment (well-formedness, truth, felicity)	Type VI—explication by paraphrase, scenario	Type VII—demonstration of referents; act-out tasks
Contact language utterance	Type II—translation	(these combinations go beyond target language elicitation)			
Content of a linguistic representation (in the target or contact language)	Type III—production in a given contextual scenario				
Content of a non-linguistic representation	Type IV—description				

a linguistic representation used as stimulus is a context description employed in combination with a stimulus utterance to test whether the utterance is considered true and pragmatically appropriate in the context.

The response in turn can be a target language utterance, a contact language utterance, a metalinguistic utterance, or some non-linguistic communicative action. Among metalinguistic responses, two types may be further distinguished: judgments and descriptions. The former rank a property of the stimulus (or some part of it), such as its acceptability, idiomaticity, unusualness, and so forth, on a scale, whereas the latter paraphrase its meaning or describe a scenario or a setting, and so forth, in which it might be used.

As Table 1.2 shows, target language utterances are the only stimuli that can be used to elicit all of the five response types. All other stimulus types yield exclusively one valid response type: a target language utterance. This means that every form of (target language) linguistic elicitation involves target language utterances as stimulus, response, or both—which makes sense. As for the empty cells, these combinations yield responses that are not valid as data for target language studies. For example, a non-linguistic stimulus might be used to elicit a contact language response in a study on the participants' second-language competence in the contact language.

The classification in Table 1.2 covers not only the methods for semantic elicitation, but—as far as I can see—those for elicitation in any field of linguistics. However, all of these methods can also play a role in semantic research, including in fieldwork. For research aiming to identify possible expressions of a given meaning, completion and association tasks, translation tasks, contextualized production tasks, and description tasks are suitable. For studies targeting the meaning of a given expression, eliciting judgments of (non-)contradiction, felicity, and so forth, explications by paraphrase or scenario, and demonstration and act-out tasks will be the methods of choice.

In the following, I provide illustrations of the seven types with examples from my own field research on Yucatec Maya. As will become clear in the process, elicitation often involves combinations of the seven types of techniques listed in Table 1.2.

4 Type I: From Target Language Utterance to Target Language Utterance

Completion and association tasks involve both a target language stimulus and a target language response. They are powerful tools for studying syntagmatic lexical relations such as selectional restrictions. I have employed association tasks in several studies of selectional restrictions. One example is a study of the semantics and argument structure of Yucatec verbs of cutting and breaking (or “separation in material integrity” (Hale and Keyser 1987; cf. Bohnemeyer 2007; Majid et al. 2008). My objective was to determine which verbs impose narrow selectional restrictions

24 Methodologies in Semantic Fieldwork

on the theme or patient and which impose such restrictions on the instrument. The hypothesis I was testing, extrapolated from Guerssel et al. (1985), was that theme/patient-specific verbs have syntactic properties similar to those of English *break*, while instrument-specific verbs have syntactic properties similar to those of *cut*. According to this hypothesis, members of the *break*-type class, but not the *cut*-type class, would produce inchoative intransitive variants that express the state change of the theme, but omit the cause, whereas members of the *cut*-type class, but not the *break*-type class, would produce variants that express pure activity meanings without a state change component. In English, these patterns are instantiated by the causative-inchoative alternation in the case of the *break*-class and the conative alternation in the case of the *cut*-class.

The procedure I used was as follows: for each verb I wanted to test, I gave the speakers I ran the study with a typical-theme prompt of the format in (7) and a typical-instrument prompt of the format in (8) (I administered the task in Yucatec, using Yucatec prompts):

- (7) “I want you to tell me the kinds of objects that can be VERBed. If you hear that somebody VERBed something, what kind of thing are you going to think it is that they VERBed?”
- (8) “I want you to tell me the kinds of objects that one can VERB with. If you hear that somebody VERBed something, what kind of thing are you going to think it is that they VERBed it with?”

I ran the task with five speakers. Tables 1.3–1.4 list the responses for *hat* “tear” and *xot* “cut”.

Cursory inspection suggests that the typical themes of *hat* “tear” form a fairly coherent set, involving objects that might be conceptualized as being made of materials of a fibrous structure. This is not actually the case for the plastic bag; but one can imagine that the category is extended to such objects as plastic bags

TABLE 1.3

Responses for *hat* “tear”

Responses to theme prompt (7)	Responses to instrument prompt (8)
Clothes, paper, leather, a plastic bag, a letter, one’s hand, one’s mouth/lips and shoes	One’s hands, feet, mouth, a stick, a machete, knife, axe, a piece of wire, scissors

TABLE 1.4

Responses for *xot* “cut”

Responses to theme prompt (7)	Responses to instrument prompt (8)
Rope, melons, squash, tomatoes, one’s hand, one’s clothes, a plank or the table, or another person	A handsaw, knife, machete, reaping hook, hacksaw, axe, shards of glass, pieces torn off an aluminum can

because “separation in material integrity” occurs in them in a manner similar to that typical of fibrous materials. In contrast, coherence in the responses to the typical-instrument prompt is fairly loose. On the basis of this observation, it might be tentatively concluded that *hat* “tear” is semantically theme-specific, but not instrument-specific. Conversely, responses to the prompts for *xot* “cut” show coherence in the instrument set (all typical instruments can be applied in the manner of bladed tools, whether or not they actually have blades), but much less so in the theme set. So this is a verb that seems more likely to be instrument-specific rather than theme-specific. For further results and analysis, see Bohnemeyer (2007). Classic readings on association include Ervin and Landar (1963) and Clark (1970). A very interesting recent application can be found in Evans and Wilkins (2000).

5 Type II: From Contact Language Utterance to Target Language Utterance or Vice Versa and Type III: From Linguistic Representation of a Stimulus Content to Target Language Utterance

A translation task directs a speaker to translate a stimulus utterance in the contact language into a response in the form of a target language utterance or vice versa.

Translation tasks are potentially fraught with two problems. First, they offer insufficient control over how the speaker construes the stimulus. For example, the speaker and the researcher may differ in their competence in the contact language or use different varieties of it, or they may differ in the inferences involved in their understanding of the stimulus utterance as a result of differences in cultural knowledge. The second potential concern is the risk of interference effects: when a speaker has a choice between two or more translations, all of which are well-formed in the target language and roughly express the intended meaning, their choice may be influenced by a desire to mimic structural properties of the stimulus. For example, in a language without definite articles, a speaker might be tempted to translate a definite article in the stimulus using a demonstrative—especially if their own imperfect understanding of the function of definite articles in the contact language treats them as equivalents of text-deictic uses of demonstratives in their native language.

Both of these pitfalls can to some extent be checked by providing the stimuli with contexts that restrict their interpretation, thereby combining Types II and III, since the context is a stimulus in its own right and a Type-III stimulus at that, the content of a linguistic representation serving as a stimulus. The most widely known and successful example of this hybrid approach is the Tense-Mood-Aspect Questionnaire of Dahl (1985).

In Dahl’s questionnaire, the translation stimuli are utterances that express event descriptions from a certain temporal, aspectual, and modal perspective. To avoid interference from the contact language, expressions of tense, aspect, and mood are omitted from the stimuli and finite verb forms are replaced with

non-finite ones (set in capital letters to flag them) wherever possible.² Instead, the intended perspective is controlled by a context that defines a reference time (or “topic time” in the sense of Klein 1994) for the translation stimulus. This is illustrated in (9). The context precedes the translation stimulus and is set in brackets.

- (9) TMA Questionnaire item A1:
 [Q: What your brother DO when we arrive, do you think?
 (= What activity will he be engaged in?)]
 He WRITE letters.

The translation target in (9) is simply a description of somebody (a male referent) writing multiple letters. The translation stimuli are designed so as to cover all major lexical-aspectual classes that have been identified across languages in previous research.³ The context defines a topic time for this description that lies in the future of the utterance time of the stimulus and is included in the runtime of the letter-writing event. In English, the future progressive is the canonical way of expressing this perspective. Example (10) is Yucatec response to the stimulus in (9). In my experience, the best way to ensure that the speaker takes the context fully into account during the translation is to ask the speaker to translate the context as well.

- (10) Q: Ba'x a=tukul-ik
 what(B3SG) A2=think-INC(B3SG)
 “What do you think?”
 k-u=beet-ik a=suku'n chéen k'uch-uk-o'n?
 IMPF-A3=do-INC(B3SG) A2=elder.brother SR.IRR arrive-SUBJ-B1PL
 “will your big brother be doing (lit. is he doing) when we arrive?”

²These infinitives tend to be confusing, however, when the task is administered purely orally, for example when working with speakers not accustomed to reading.

³A fundamental problem for any research that starts from a set of semantic or notional categories and asks how these are expressed in a given language is what is known in the typological and ethnosemantic literature as the ‘etic grid’ problem: the set of semantic or notional categories that the study is designed to test—the study’s etic grid—biases the possible observations of semantic categories in the target language. Nowhere has this issue been discussed more prominently—or more pointedly—than in the critique advanced by Lucy (1997), Saunders & van Brakel (1997), and others of the methodology of Berlin & Kay’s (1969) seminal study of the semantics of color terminologies (see Berlin & Kay (1997) and Kay (2006) for the authors’ response). Studies that aim to examine the expression of a given semantic domain in particular languages should base their etic grid on a careful review of the distinctions reported in the available crosslinguistic literature; should always be mindful of the fact that the sets of categories they encounter in grid-based elicitation are partly a function of the grid they employ; should strive to compare the results of the grid-based elicitation to data obtained from other (especially non-elicited) sources (see Bohnemeyer 2012 for an illustration); and should present their findings as an initial step in a series of studies, to be followed up by further research based on a revised grid that takes into account any shortcomings of the initial grid that emerged during the first round or from the comparison with other sources of evidence. See also AnderBois & Henderson, this volume.

- A: Chéen k'uch-uk-o'n wal=e',
 SR.IRR arrive-SUBJ-B1PL UNCERT=D3
 "When we arrive, I guess"
 ts'ïib-t-ah+kàartah k-u=meet-ik wal=e'.
 write-APP-ATP+letter(B3SG) IMPF-A3=do-INC(B3SG) UNCERT=D3
 "letter writing is what he will be (lit. is) doing, I guess."

The response features the use of the irrealis marker *kéen/chéen*, which is restricted to subordinate clauses and governs subjunctive mood; the imperfective aspect marker in the matrix clauses; a predicate focus construction; and the epistemic uncertainty particle *wal*. None of these expressions encodes relative or absolute future tense.

6 Type IV: From Non-Linguistic Representation to Target Language Utterance

The elicitation of descriptions of non-linguistic stimuli has become the method of choice in semantic typology since the landmark study by Berlin and Kay 1969 (with much earlier precursors such as Magnus (1877, 1880) in research on the linguistic categorization of color and Chamberlain 1903 and Myers 1904 on that of tastes). It also plays a prominent role in speech production research and language acquisition research. Moreover, non-linguistic stimuli are not only used in production tasks, but also in various types of comprehension tasks and in so-called "referential communication" tasks (see below), which combine production and comprehension. Either way, the principal function of non-linguistic stimuli is to constrain the referential content of a target language utterance—the response in the case of production tasks and a second stimulus, a stimulus utterance, in the case of comprehension tasks. One important caveat for production tasks with non-linguistic stimuli is that *constrain* does not mean the stimulus fully determines the meaning of the response. The meaning of the speaker's response will depend above all on their interpretation of both the stimulus and the task, or their interpretation of the researcher's intention behind both. Let me illustrate the role of the task first. Consider Figure 1.4, which shows the first item in the *Topological Relations Picture Series*, a.k.a. BowPed. BowPed consists of 71 line drawings featuring spatial configurations. Most of these involve "topological" relations in the sense of Piaget and Inhelder (1956), (i.e., relations that can be adequately described without selection of a perspective or reference frame). In each picture, one or more objects are designated as "figures" (Talmy 2000) or themes of locative descriptions by arrows pointing to them. The participants' task is to use the information in the picture to answer the question "Where is the [figure]?" asked preferably in the target language. This question serves as secondary stimulus, thus making BowPed strictly speaking a combination of Types III and IV.

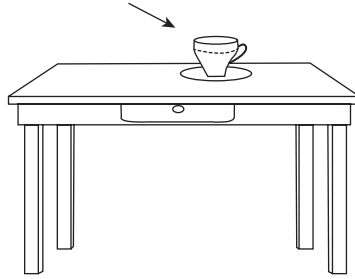


FIGURE 1.4 Item #1 of the “Topological Relations Picture Series” aka BowPed (©Eric Pederson; reproduced with permission)

However, when I ran BowPed with Yucatec speakers, I had an experience I have heard several other researchers report: in response to the question “Where is the cup?,” a speaker would look at me with mild puzzlement and point to the picture: “Uh, right here?” What this response brought home is that I had not been specific enough about the task. To fix this problem, I constructed the following scenario, which I asked the speaker to assume as an *elicitation frame*, that is, a more elaborate context within which to respond, a context that put a certain interpretation on the *Where*-question:

- (10) “Imagine you are talking to somebody who is looking for the [figure]. This person knows where the [ground] is, but does not know where the [figure] is. You know where the [figure] is; but neither of you can see the [figure] and the [ground] right now. The person asks you “Where is the [figure]?” Imagine you want to tell the person where the [figure] is. How do you respond?”

I would repeat this frame with every new picture until I got the impression that the speaker understood and remembered the point.

In comprehension tasks, the visual stimulus is presented along with a target language utterance—a typical example of a hybrid technique. For instance, in verification tasks, the speaker is to determine whether the utterance can serve as a description of the visual stimulus (or, more generally, whether it instantiates its extension—a type of judgment elicitation). In matching tasks, another subtype of comprehension tasks, the speaker is asked to select from among two or more visual representations the one that is best (most accurately, etc.) described by the utterance, or select among two or more utterances the one that best describes a given visual representation. Verification and matching tasks thus combine elements of Type IV and Type V. Referential communication tasks are a combination of production and comprehension distributed across two participants (cf. Clark and Wilkes-Gibbs 1990). They involve at least two speakers per trial: one describes the content of a stimulus and the other rematches the description to a set of non-linguistic stimuli. There are numerous possible realizations of this, including picture to picture, picture to toy model, and so forth. Figure 1.5 illustrates the setup of a picture-to-picture matching task.



FIGURE 1.5 Setup of the Ball & Chair picture matching task

The “Ball & Chair” referential communication task (Bohnmeyer 2011) was developed to replace and improve upon a similar task, “Men & Tree,” designed at the Max Planck Institute for Psycholinguistics in the 1990s (Pederson et al. 1998). The goal of both tasks is to assess the participants’ use of spatial frames of reference in discourses referring to small-scale space. To this end, two speakers sitting side by side are asked to match identical sets of photographs placed in front of them in different orders, while a screen between them prevents them from sharing a visual field (I used the suitcase in which I had hauled my field equipment). The screen forces the participants to produce maximally explicit descriptions in order to solve



FIGURE 1.6 Ball & Chair picture 2.5

30 Methodologies in Semantic Fieldwork

the task. The photos all show a ball and a chair. There are four sets, each comprising 12 pictures, which differ from one another in the orientation of the chair and the location of the ball vis-à-vis the chair. Example (11) reproduces a description of one picture, shown in Figure 1.6, in full.

- (11) a. Estée, u séegere-e-e . . . chan fòotoa',
 esté u=séegir le=chan fòoto=a',
 HESIT A₃=follow DET=DIM photo=D₁
 "Uh, this next-uh-little photo,"
- b. u frèente e siiyaó, tu tohile don Jorgeó,
 u=frèente le=siiya=o' tu=tohil le=don Jorge=o'
 A₃=front DET=chair=D₂ PREP:A₃=straight:REL DET=don Jorge=D₂
 "the front of the chair, in the line of that don Jorge (i.e., JB),"
- c. ti' yàani'. Tu'x ku nakta' máako',
 ti'=yàan=i' tu'x k-u=nak-tal máak=o'
 PREP=EXIST(B₃SG)=D₄ where IMPF-A₃=lean-INCH.DIS person=D₂
 "there it is. The back rest (lit. where a person leans (against)),"
- d. estée, ta frèente súutu'.
 estée ta=frèente súut-ul
 HESIT PREP:A₂=front turn\MIDDLE-INC(B₃SG)
 "uh, it's turned (toward) your front."
- e. Ta xno'hk'abile'
 ta=x-no'h+k'ab-il=e'
 PREP:A₂=F-right+hand-REL=TOP
 "On your right,"
- f. ti' yàan ump'ée bòolai'.
 ti'=yàan hun-p'èel bòola=i'
 PREP=EXIST(B₃SG) one-CL.IN ball=D₄
 "there is a ball."
- g. Ta xts'ii - ta xts'iihk'abil [unintel.], ti' yàan
 ta=x-ts'ii+k'ab-il=e' ti'=yàan
 PREP:A₂=F-left+hand-REL=TOP PREP=EXIST(B₃SG)
 ump'èel bòolai',
 hun-p'èel bòola=i'
 one-CL.INball=D₄
 "On your le—on your left [unintelligible], there is a ball,"

preferences across participants, one can assess the preferences of a generic or average or typical Maya speaker in rural central Quintana Roo. This in turn permits comparisons across speech communities both among speakers of the same language and among speakers of different languages. It allows us to conclude, for example, that rural Yucatec speakers make more frequent use of relative frames in solving this artificial task than speakers of many other Mesoamerican languages, but do so much less frequently than speakers of European languages (see the descriptions in O'Meara and Pérez Báez (eds.) 2011, including Bohnemeyer 2011). This is an interesting and important finding: it suggests that relative reference frames cannot play the same role in reference to small-scale space that they play in Euro-American speech communities (and, e.g., among Japanese speakers (Kita 2006)), where they are the default for this domain. But it does not tell us much about what Yucatec speakers habitually do to communicate about space. Assuming that natural referential practice relies heavily on gaze and gesture, it is a foregone conclusion that designs such as Ball & Chair necessarily produce rather distorted representations of it. The standard response in the semantic typology community to this problem has long been that elicitation results—especially, but not restricted to, results obtained with referential communication designs—should always be complemented by other sources of evidence, both elicited—for example, in the case of spatial reference, route descriptions—and non-elicited, that is, staged discourses (firsthand witness accounts of natural disasters and local history narratives may prove useful in spatial studies) and the observation of spontaneously occurring interactions. More on this below.

When it comes to elicitation stimuli, considerations of ecological validity must take into account three factors:

- ❑ Are there conventional descriptors for the stimuli in the target language?
- ❑ Are the stimuli culturally appropriate?
- ❑ How do speakers of the target language interpret the stimuli?

The first issue is usually the most trivial in my experience, as it tends to be confined to lexical expressions. At the lexical level, the problem of missing descriptors is readily addressed by the researcher negotiating with the speakers either the use of a contact language loan or a reinterpretation of the stimulus item in question that makes it describable in the target language. For example, if a stimulus picture or video shows a plant or animal of a species that does not occur in the local environment, it may be possible to ask the speakers to treat it as an instance of a similar plant or animal that does occur. Lexicalization gaps can create problems that spill over into the grammar in cases in which the grammatical classification of lexical items is concerned, for example when noun class or *aktionsart* markers are involved.

As for the second issue, both linguistic and non-linguistic stimuli can be offensive to members of particular cultures for a variety of reasons: exposure of body parts that is considered indecent, characters hunting animals or eating foods

considered taboo, and so forth. There is no other solution to this type of problem than to avoid it during the design of the stimuli.

Lastly, the interpretation of visual stimuli is subject to non-trivial cultural conventions. Consider Figure 1.7, a line drawing created by David Wilkins as part of a series of stimuli designed for the elicitation of expressions of manner of motion by children learning the Pama-Nyungan language Arrernte spoken in and around Alice Springs in central Australia (see Wilkins 1997). The intended interpretation of the picture was that of a horse in full gallop. However, the Arrernte children instead understood it as showing a dead horse lying in the dirt. What was intended to be seen as clouds of dust thrown up into the air by the horse's legs was instead understood as the traces the onset of rigor mortis had left behind in the sand. These different interpretations are the result of different cultural conventions for visual representations: whereas the default perspective for such representations in Asian and European cultures is horizontal, it is the bird's-eye view in Aboriginal cultures. The different conventions in turn may be linked to the most widespread traditional media for visual representations in each culture: paper and canvas in Eurasian cultures vs. campground dirt in cultures of Aboriginal Australia.⁴ (Now, of course, all of these materials are increasingly being replaced by digital media, with the inevitable result of a globalization of the horizontal perspective.)

The dependence of stimuli on culture-specific interpretations only increases with the semiotic complexity of the stimuli. Consider, for example, the representation of events by single snapshot images vs. cartoon-strip sequences vs. video clips. This, too, is subject to changing cultural conventions—e.g., medieval and non-western artists often represent temporal as spatial relations, as in the case of the

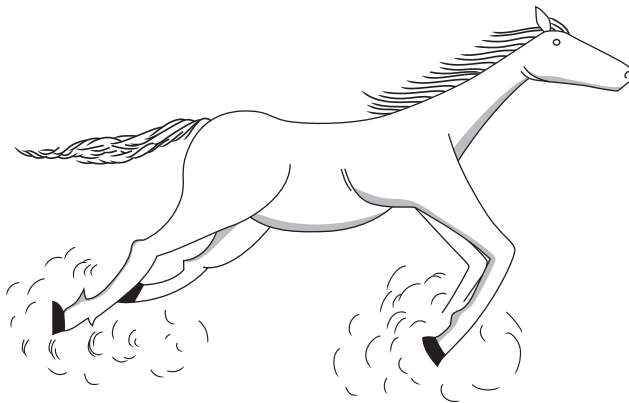


FIGURE 1.7 Galloping horse or dead horse? (Wilkins 1997: 157; ©David P. Wilkins; reproduced with permission)

⁴Sand drawings on the campground are of course exclusively viewed from above—hence the naturalness of the bird's-eye perspective.

Bayeux Tapestry, which shows the events of the Norman conquest of England as if they all happened simultaneously but in adjacent places. In contrast, contemporary Western imagery strictly follows a convention according to which everything that is represented within the same drawing is understood to (have) happen(ed) simultaneously. Consequently, representing a sequence of events requires a sequence of images, for example in separate panels, as in a comic strip.

Guarding against the effects of culture-specific interpretations of visual stimuli is but one example of a much more general principle: a stimulus only impacts the response via the speaker's interpretation of it. This is captured by the Golden Rule of Elicitation proposed below.

Another often-commented-on limitation of non-linguistic stimuli is their restriction to perceivable and thus concrete information. This limitation can be overcome by combining non-linguistic and linguistic stimuli or through complex task designs. An example is the TEMPEST design for the elicitation of temporal relations described in Bohnemeyer (1998a, b; 2000), a referential communication task in which speakers match videos that show the same events in contrasting orders. Cf. also Burton and Matthewson (this volume) and Bar-el (this volume).

7 Type V: From Target Language Utterance to Judgment

Judgments are metalinguistic utterances that may comment on a variety of properties of linguistic stimuli: their grammaticality, interpretability, idiomaticity, stereotypicality, pragmatic appropriateness (which covers a large variety of different properties; see above), and—arguably most importantly for the purposes of empirical semantics—whether particular individuals or states of affairs are elements of their extension. These metalinguistic utterances are typically prompted by questions or requests. The cognitive basis of such judgments is not entirely understood, but is presumably related to the speaker's ability to detect ill-formed, semantically false, or contextually inappropriate (constituents of) utterances both in processing the speech of their interlocutors and in their own production.

As pointed out by Matthewson (2004) and Tonhauser et al (2013), it makes little sense to ask a linguistically untrained speaker whether an utterance has a given entailment, carries a certain presupposition, and so forth. This is trivially true for the simple reason that untrained speakers will not have the relevant technical notions of “entailment,” “presupposition,” and so on. This means that it is up to the skill of the researcher to construct an elicitation stimulus and task that allow the speaker to express *a* judgment from which the researcher can then infer whether or not the utterance has the relevant property in the speaker's judgment. The successful elicitation design must circumvent the problematic technical notions by instead tapping into the definition of the relevant semantic property.

For illustration, consider entailment. An utterance entails another if any possible world that makes the former true also makes the latter true. To test whether

this is the case for a given pair of utterances, the researcher may ask the speaker whether they can think of a situation in which the first utterance is true but the second is not. This question is likely still too complex and abstract for most untrained speakers to readily answer it on first trial. To simplify matters further, the researcher can construct a few scenarios on a trial basis and ask the speaker whether both utterances are true in them. This of course means that the researcher is not asking for a direct judgment of entailment, but rather for a series of judgments about the truth of a pair of utterances in a series of scenarios. Another important avenue for eliciting entailment data are judgments of contradiction. Speakers appear to be able to tell relatively immediately whether two statements are logically consistent or not. Consequently, one method for testing whether an utterance has a given entailment is by combining it with a second utterance, which negates the hypothetical entailment. If in the speaker's judgment the conjunction of the two utterances may be true in the same scenario, this suggests that the proposition negated by the second utterance is not an entailment of the first. But if the speaker judges the utterances to be inconsistent, this supports the entailment analysis.

Judgments are almost always of a graded nature. That is, even if a speaker gives a categorical response to a simple polar question, this response can be ranked in relative strength with respect to the same speaker's responses to other stimuli.

There are a number of principal obstacles that may beset the elicitation of judgments:

- ❑ Judgments may not reflect the speaker's own production well.
- ❑ Judgments may be influenced by normative beliefs.
- ❑ The same stimulus utterance may be judged differently by the same speaker in different contexts.
- ❑ A speaker's judgment in response to a particular utterance will depend on which aspect of the utterance the speaker understands they are asked to judge, or in other words, which type of judgment they are asked to make—a judgment of well-formedness, idiomaticity, and so forth. However, linguistically untrained speakers may not find it easy to distinguish between these different types of properties and judgments.
- ❑ Similarly, a linguistically untrained speaker cannot always be expected to be able to locate the source of a violation of well-formedness, idiomaticity, interpretability, or the like. In general, a speaker can tell that an utterance "sounds funny" (in a given context), and may even associate the anomaly with non-native speakers of a particular background. But they are less likely to be clear on why the utterance "sounds funny."

I have encountered the phenomenon alluded to in the first point on numerous occasions: speakers will reject a certain construction or the use of a certain term in reference to a particular state of affairs and later produce that very construction during a different task or use that very term in reference to the state of affairs.

There can be a variety of reasons for why a decontextualized expression appears to us differently than when we come upon a context in which that same expression is used by others or in which we might use the expression ourselves. Moreover, judgments are always susceptible to normative beliefs—the second point above—and such beliefs may cause speakers to reject particular expressions even though they themselves use them. Such beliefs may not always be the result of standardization, but can also be influenced by folk theories of language use. As an example, my work on spatial reference frames mentioned in the previous section has taught me that many speakers of Yucatec and other Mayan languages operate on a belief that tasks such as the one shown in Figure 1.4 have correct and incorrect solutions and that the correct ones employ cardinal direction terms. The origin of this belief is at present unclear to me.

An instance of the context-dependence of judgments that many linguists are familiar with from their own practice and that is also well documented in the psycholinguistic literature is that of *satiation*: the phenomenon that the acceptability of utterances that appear initially anomalous sometimes seems to improve with time as the same utterance is repeated again and again (Snyder 2000; Hiramatsu 2000; Goodall 2005; Francom 2009; *inter alia*). Some types of anomaly are known to satiate much more easily than others. Why this is the case is unknown, and the causes of satiation itself are poorly understood.

The ability to distinguish between different types of anomaly—as induced by syntactic vs. semantic vs. pragmatic clashes, and so forth.—depends on a consultant's declarative, metalinguistic understanding of linguistic phenomena and thus grows with the consultant's experience and training (see also Cover and Tonhauser, this volume). An independent potential challenge may be the terminology available to the researcher and the native speaker consultants—in either the target language or a contact language—to distinguish the relevant sources. My preference has always been to ask general acceptability questions and try to construct the stimuli so as to minimize the risk of ambiguity in the speaker's response. Researchers should of course always aim to make sure that their stimuli do not feature any anomalies other than the one to be tested. But they cannot possibly always succeed at this unless they are omniscient about the target language except perhaps for the anomaly under investigation. Typical query formats I would use are listed in (12):

- (12) a. "What about this one, how does it sound to you: [stimulus utterance]"
- b. "[stimulus utterance] Is it said well like that?"
- c. "[stimulus utterance] Is it possible to be said like that?"
- d. "[stimulus utterance] Are there people, you think, who talk (lit. say it) like this?"
- e. "In the photo/picture/video here, can it be said that [stimulus utterance]?"
- f. "In the photo/picture/video here, if a person says that [stimulus utterance], would that be true?"

- g. “Let’s say [verbal description of scenario]. In that case, can it be said that [stimulus utterance]?”
- h. “Let’s say [verbal description of scenario]. In that case, if a person says that [stimulus utterance], would that be (lit. is that) true?”

The templates in (12a–d) might be used to test the well-formedness of an utterance as per its morphosyntactic and morphophonological structure and the selectional restrictions of its lexical items. In contrast, (12e–h) can be used to test whether a given description is accurate and pragmatically appropriate in reference to a particular nonverbal stimulus (12e–f) or a verbally described scenario (12g–h). Both of these options are illustrated below.

To test whether a particular entity or state of affairs falls into the semantic extension of a given descriptor, the researcher can ask speakers, in the simplest case, “Can *X* be called *Y*?” where *X* is a verbal or nonverbal representation of the referent or simply an instance of it and *Y* is the descriptor to be tested. However, unless the descriptor is a non-relational common noun—and often even then—it is generally preferable to insert it into a declarative sentence and study its impact on the truth conditions of assertions of such a sentence. This can be understood as an application of the Context Principle often attributed to Frege 1884 and Wittgenstein 1921.

The referent *X* to be tested for inclusion in the extension of the expression can be an actual instance of a particular kind of entity or state of affairs or a representation of it. The representation in turn can be linguistic or non-linguistic. Let me illustrate elicitation of truth judgments against both verbally and nonverbally represented scenarios, beginning with the former type. My example for this comes from testing Yucatec verb phrases for telicity. Telicity has no syntactic reflexes in this language (Bohnmeyer 2002: 172–192). There is, for example, no distinction between duration (i.e., *for*-type) and time-span (i.e., *in*-type) adverbials. “Spend *X* time VERBing” and “take *X* time to VERB” are expressed the same way. The aspectual verbs translating “finish”/“complete” are compatible with telic and atelic verb phrases alike.

The only way to test for telicity is by tapping into the entailment patterns collectively known as the “imperfective paradox” (Dowty 1979; cf. also Bohnmeyer and Swift 2004), which are used in Vendler 1957 to distinguish accomplishments from activities and thus effectively to define telicity (even though Vendler did not use that notion). As illustrated in (13), activity descriptions in the progressive entail their simple-tense counterparts (13a), whereas the same does not hold for accomplishment descriptions in the progressive (13b):

- (13) a. *Floyd was pushing a cart*
 ∴ *Floyd pushed a cart.*
- b. *Floyd was drawing a circle.*
 not ∴ *Floyd drew a circle.*

To study the behavior of particular Yucatec verbs vis-à-vis these entailment patterns, I consulted with five native speaker consultants to find scenarios in which the event described by the verb phrase is plausibly interrupted at a time at which the VP marked for progressive aspect applies. I then asked whether a perfective or perfect form of the same VP can be truthfully asserted at the time of the interruption.

- (14) Pedro=e' táan u=k'àay,
 Pedro=TOP PROG A₃=sing\ATP
 "Pedro, he was singing,"

káa=t-u=k'at-ah u=báah Pablo.
 CON=PRV-A₃=cross-CMP(B₃SG) A₃=self Pablo
 "(when/and then) Pablo interfered."

Pedro=e' t-u=p'at-ah u=k'àay.
 Pedro=TOP PRV-A₃=leave-CMP(B₃SG) A₃=sing\ATP
 "Pedro," he stopped singing."

Be'òora=a' ts'ò'k=wáah u=k'àay Pedro?
 now=D₂ TERM=ALT A₃=sing\ATP Pedro
 "Now, has Pedro sung?"

The researcher should be prepared for surprises: for example, most consultants answer negatively in response to (14) since *kàay* "sing," the antipassive stem of the transitive root *k'ay* "sing," is normally interpreted as "sing a song" (cf. Bohne-meyer 2002:172–199 for details).

If possible, a visual stimulus should be used to clarify the scenario against which one wishes to test entailments. This is the verification method mentioned above. As an example, in a study reported in Bohne-meyer 2010, I examined whether Yucatec verbs of "inherently directed motion" (Levin 1993) entail translational motion of the figure or merely change of location, as described by Kita 1999 for Japanese *hairu* "enter" and *deru* "exit." To test this, I employed the *Motion verb* (*MoVerbs*; Levinson 2001). *MoVerbs* comprises 96 computer-animated video clips featuring a variety of location change scenarios varied according to the spatial relation between the "figure" or theme and some reference entity or "ground" in the source or target state or in between, the involvement of figure motion, and the perspective (toward/away from observer vs. lateral to the observer's viewing axis). I would, for example, test whether Yucatec speakers find (15) acceptable in reference to the clip whose first and last frame are depicted in Figure 1.8, in which a plank slides under a ball and cylinder:

- (15) H-na'k le=chan kaniika y=óok'ol le=tàabla=o'
 PRV-ascend(B₃SG) DET=DIM marble A₃=on DET=plank=D₂
 "The marble, it went up the plank"

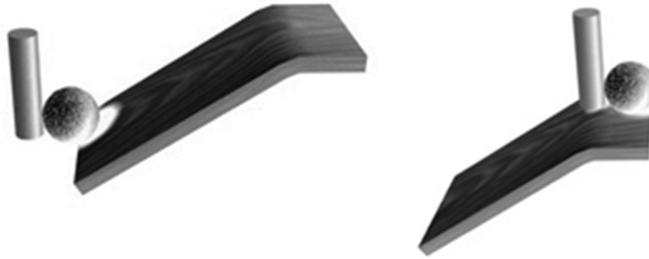


FIGURE 1.8 First and last frame of “FIGURE_GROUND 14” (Levinson 2001; ©Stephen C. Levinson; reproduced with permission)

Out of context, Yucatec speakers will reject this description in reference to the scenario shown in the clip. However, as discussed in the next section, this is not because the scene violates an entailment of (15), but rather because it is incompatible with a stereotypical interpretation of it, which a hearer will assume by implicature unless it is canceled or blocked.

8 Type VI: From Target Language Utterance to Linguistic Representation

A very powerful strategy for elucidating the meanings of linguistic expressions involves a reversal of sorts of the method described in the previous section: ask speakers to come up with and describe to you a scenario in which a given utterance might be used to make a truthful and pragmatically appropriate statement. For example, I asked speakers to modify (15) above to turn it into an acceptable description of the scenario depicted in Figure 1.8. One response to this procedure is shown in (16):

(16) Le=chan t̄aabl̄a=o' h=p̄eek-nah-ih, k̄aa=h-na'k
 DET=DIM plank=D2 PRV=move-CMP-B3SG CON=PRV-ascend(B3SG)

le=chan kaniika y=éetel che' te'l y=óokol=o'.
 DET=DIM marble A.3=with wood there A3=on=D2

“The little plank, it moved, and the little marble and the tree ascended there on top.”

As soon as it is made explicit that it was the ground—the plank—that moved, this speaker has no problem with asserting the location change description of the stationary figure. This strongly suggests that the inference to figure motion triggered by (15), which clashes with the scenario shown in Figure 1.8, is merely an implicature, not an entailment. The most likely type of implicature is a stereotype implicature—a generalized conversational implicature licensed by Grice’s second

Quantity maxim: for Yucatec speakers, just like for English speakers, the stereotypical way for someone or something to change location is for them/it to move.

“Reverse-engineering” scenarios or contexts in which a given expression might be used can provide powerful insights into the semantics and pragmatics of the expression. However, not every speaker will find the task of coming up with such an instantiation equally easy to solve. It requires imagination, a gift apparently not evenly distributed among people. In my experience, of all the skills that may qualify a good native speaker consultant, the ability to envision scenarios and contexts is the rarest and most precious for the purposes of empirical semantic research.

9 Type VII: From Target Language Utterance to Non-Linguistic Representation

The final type of elicitation task has speakers produce a non-linguistic representation of the meaning of a target language expression. *Demonstration* seems to me an appropriate general label for this type of task. A special subtype of demonstrations are act-out tasks, in which a speaker instantiates a described action or event literally or by playacting it (or through a combination of both).

The example I would like to offer to illustrate demonstrations as a type comes from a study I did a few years ago on the semantics of “dispositional” roots. Such roots lexicalize non-inherent spatial properties such as postures. Mayan languages have hundreds of roots of this kind, and the majority of these select for inanimate figures. For this reason, I prefer “dispositional” to the traditional Mayanist term “positional,” which suggests postures. In Yucatec and many other Mayan languages, dispositionals represent a root class *sui generis* with unique privileges of producing stems of various lexical categories, among which verbs do not necessarily stand out. Yucatec dispositional roots produce transitive and intransitive verb stems, derived stative predicates, numeral classifiers, and more, depending on the derivational morphology used (Bohnemeyer and Brown 2007). Distinctions that enter the conceptualization of dispositions include support, suspension, blockage of motion, orientation (mainly in the gravitational field), shape, and configuration of parts of the figure with respect to one another. Location is not a dispositional concept; rather, dispositions can be thought of as “manners of location” (Belloro et al. 2008).

The greatest challenge in analyzing dispositional semantics is that the dimensions of contrast are poorly understood, since dispositions are not lexicalized in Indo-European languages at the level of specificity at which they are lexicalized in Mayan languages. To overcome this challenge, I applied a two-step process inspired by Berlin’s 1968 classic study of Tzeltal numeral classifiers. In a first step, I elicited typical themes or figures for each previously identified dispositional root with six speakers, applying an association task very similar to the typical-theme and typical-instrument prompts described above. I then consolidated the responses by

identifying the 20 most frequently mentioned types of themes—humans; various species of animals, including horses, dogs, birds, and snakes; ropes; clothes and pieces of fabric; and so forth—and the total set of roots in association with which each type of figure had been mentioned by at least one speaker. Then, in a second elicitation phase, I asked the same six speakers to demonstrate all the dispositions associated with a given type of figure contrastively, by showing me how it would have to be manipulated to get it from an instantiation of the last demonstrated disposition to one of the disposition described by the root I was prompting the speaker with now. For some of the figure types, actual exemplars were used; others were represented by toys. I videotaped these sessions, resulting in a total of about 26 hours of videotape. Since then, several students in the University at Buffalo Semantic Typology Lab have been working on the coding of these video files, attempting to identify the properties shared across demonstrations of the same root-figure pair by different speakers and those that distinguish dispositions expressed by different roots. Figure 1.9 illustrates four types of suspension configurations described by different roots. This work is generating hypotheses regarding the semantics of the roots, which are being tested in follow-up fieldwork.

The Dialectical Pivot: Empirical and Hermeneutic Approaches Revisited

This chapter started from the premise that semantic research within the social and behavioral sciences must be an empirical endeavor based on the observation of



FIGURE 1.9 Suspension dispositions described by (clockwise from top left) *choh*, *ch'uy*, *lech*, and *t'oy*

the communicative behavior of interlocutors. Yet, somewhat paradoxically, having taken the reader several steps along the way toward an answer to the question how empirical semantic research is possible, I am now about to argue that a mature empirical semantics must in fact avail itself of the techniques of hermeneutic analysis in order to achieve its goal.

We have seen above that a speaker's response to an elicitation stimulus depends on the speaker's interpretation of the stimulus. This of course holds not just for linguistic elicitation, but for any type of empirical research with human or animal participants. It is a valid question—and one routinely asked—in the analysis of experimental results in psychology or interview responses in sociology and political science how the presentation of the task and/or the stimuli may have influenced the observed responses. Think, for example, of how easily the findings of a marketing research study or an opinion poll can be influenced by the way the questions are asked, the orders in which they are asked, and of course the sampling procedures used to recruit participants. These are all questions that go to the validity of study designs in empirical research. But linguistic elicitation adds a potential further layer to this problem complex. In any *quantitative* research design, the goal is to determine whether there are significant correlations between predictor and response variables. As long as the research design is valid, any such correlation is a reportable outcome, and so is the absence of a predicted correlation. Linguistic elicitation, however, including semantic elicitation, may produce data for quantitative or qualitative analysis or both. All of the studies discussed above produced primarily data for qualitative semantic analyses, meaning analyses that draw direct conclusions concerning the meanings of particular utterances and expressions. In such analyses, the speaker's interpretation of the stimuli and task and the intended interpretation of their response—and the researcher's assumptions about all of these—are not merely a validity concern, but have direct bearing on the content of the analyses. By way of illustration, the Yucatec speakers who rejected (15) as a description of the animation in Figure 1.8 did not apparently intend this judgment to be understood to the effect that (15) would be false as a representation of the scene, but rather to the effect that it would be misleading. This, however, did not become apparent until I asked them to think about how the utterance might be amended in order to make it acceptable in reference to Figure 1.8 with the result shown in (16).

The Golden Rule of elicitation states that an elicitation response only becomes a data point in formulating generalizations about the linguistic competence and practices of language use of the members of a speech community once the speakers' interpretation of the task and stimulus and the intended interpretation of the response have been ascertained. A "raw" elicitation response does not document much of anything about the speaker's knowledge except for the fact that they are able to produce it, which does not even tell us whether the responses are well-formed.⁵

⁵A similar view is stated in Matthewson (2004).

10 Summary

Researchers who study semantics in the field working with speakers of understudied languages or in the lab working with small children have to proceed without being able to rely on their own native speaker intuitions or on those of expert speakers with linguistic training. This chapter has argued that this is not only possible, but that in fact all semantic research conceived of as part of the social and behavioral sciences should not content itself with the researcher's own native speaker intuitions as the sole source of evidence. Such introspective approaches presuppose a hermeneutic view of semantic research with interpretation as the fundamental source of evidence. In contrast, the present chapter has advocated for an empirical semantics based on the observation of communicative behavior as it reveals the referential extension of linguistic expressions, their selectional restrictions, the structure of their sense spectra, the pragmatic conditions of using them, and their processing properties. The empirical semanticist infers these properties from observations of how competent speakers use the expressions under study, not unlike a child acquiring the semantic systems of the languages she is exposed to by observing competent speakers in the act of using them.⁶

References

- Belloro, Valeria, Jürgen Bohnemeyer, Dedre Gentner, and Kathleen Braun. 2008. *Thinking-for-Speaking: Evidencia a Partir de la Codificación de Disposiciones Espaciales en Español y Yucateco* [Thinking for Speaking: Evidence from the Encoding of Spatial Dispositions in Spanish and Yucatec]. Memoria del IX Encuentro Internacional De Lingüística En El Noroeste. Vol. 2. Hermosillo: Editorial UniSon. 175–190.
- Berlin, Brent. 1968. *Tzeltal Numeral Classifiers: A Study in Ethnographic Semantics. Janua Linguarum* (All Series), Bloomington, IN; 70. The Hague: Mouton de Gruyter.
- Berlin, Brent, and Paul Kay. 1969. *Basic Color Terms*. Berkeley: University of California Press.
- Bohnemeyer, Jürgen. 1998a. Temporal reference from a Radical Pragmatics perspective: Why Yucatec does not need to express 'after' and 'before.' *Cognitive Linguistics* 9(3): 239–282.
- Bohnemeyer, Jürgen. 1998b. Time relations in discourse: Evidence from Yucatek Maya. PhD diss., Tilburg University.
- Bohnemeyer, Jürgen. 2000. Event order in language and cognition. In *Linguistics in the Netherlands 17*, H. de Hoop and T. van der Wouden (eds.), 1–16. Amsterdam: John Benjamins.
- Bohnemeyer, Jürgen. 2002. *The Grammar of Time Reference in Yucatek Maya*. Munich: LINCOM.

⁶Of course, the researcher's goal is fundamentally different from the child's: one is aiming for (primarily) declarative knowledge, the other for procedural knowledge. The approaches the two take are tailored toward these different goals and are consequently not interchangeable. What they have in common, however, is that they both rely on the same types of observational evidence.

44 Methodologies in Semantic Fieldwork

- Bohnenmeyer, Jürgen. 2007. Morpholexical transparency and the argument structure of verbs of cutting and breaking. *Cognitive Linguistics* 18 (2): 153–177.
- Bohnenmeyer, Jürgen. 2010. The language-specificity of Conceptual Structure: Path, Fictive Motion, and time relations. In *Words and the World*, Barbara Malt and Philip Wolff (eds.), 111–137. Oxford: Oxford University Press.
- Bohnenmeyer, Jürgen. 2011. Spatial frames of reference in Yucatec: Referential promiscuity and task-specificity. *Language Sciences* 33: 892–914.
- Bohnenmeyer, Jürgen. 2012. Yucatec demonstratives in interaction: spontaneous vs. elicited data. In *Practical theories and empirical practice*, Andrea C. Schalley (ed.), 99–128. Amsterdam: John Benjamins.
- Bohnenmeyer, Jürgen, and Penelope Brown. 2007. Standing divided: Dispositional verbs and locative predications in two Mayan languages. *Linguistics* 45 (5–6): 1105–1151.
- Bohnenmeyer, Jürgen, and Mary D. Swift. 2004. Event realization and default aspect. *Linguistics and Philosophy* 27 (3): 263–296.
- Bowerman, Melissa, and Eric Pedersen no date. Cross-linguistic perspectives on topological spatial relationships. Unpublished manuscript, Max Planck Institute for Psycholinguistics.
- Brown, Roger. 1958. *Words and Things*. New York: Free Press.
- Bueno, Steve, and Cheryl French-Mestre. 2008. The activation of semantic memory: Effects of prime exposure, prime-target relationship, and task demands. *Memory & Cognition* 36(4): 882–898.
- Chamberlain, Alexander Francis. 1903. Primitive taste words. *The American Journal of Psychology* 14: 146–153.
- Clark, Herbert H. 1970. Word associations and linguistic theory. In *New Horizons in Linguistics*, John Lyons (ed.), 271–286. Baltimore, MD: Penguin.
- Clark, Herbert H., and Deanna Wilkes-Gibbs. 1990. Referring as a collaborative process. In *Intentions in communication*, Philip Cohen, Jerry Morgan, and Martha Pollack (eds.), 463–493. Cambridge, MA: MIT Press.
- Costa, Albert, F.-Xavier Alario, and Alfonso Caramazza. 2005. On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review* 12(1): 125–131.
- Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.
- Dowty, David R. 1979. *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and Montague's PTQ*. Dordrecht: Reidel.
- Ervin, Susan M., and Herbert Landar. 1963. Navajo Word Association. *American Journal of Psychology* 76: 49–57.
- Evans, Nicolas, and David P. Wilkins. 2000. In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language* 76 (3): 546–592.
- Ferrand, Ludovic, and Boris New. 2003. Semantic and associative priming in the mental lexicon. In *The mental lexicon*, Patrick Bonin (ed.), 25–43. New York: Nova Science Publishers.
- van Fraassen, Bas. 1980. *The Scientific Image*. Oxford: Oxford University Press.
- Francom, Jerrid. 2009. Experimental Syntax: Exploring the effect of repeated exposure to anomalous syntactic structure—evidence from rating and reading tasks. PhD diss., University of Arizona.
- Frege, Gottlob. 1980 [1884]. *The Foundations of Arithmetic*. Trans. J. L. Austin. Second Revised Edition. Evanston, IL: Northwestern University Press.

- Goodall, Grant. 2005. On the syntax and processing of wh-questions in Spanish. In *WCCFL 23 Proceedings*, B. Schmeiser, V. Chand, A. Kelleher, and A. Rodriguez (eds.), 101–114. Somerville, MA: Cascadilla Press.
- Guerssel, Mohamed, Kenneth Hale, Mary Laughren, Beth Levin, and Josie White Eagle. 1985. A cross-linguistic study of transitivity alternations. In *Papers from the Parasession on Causatives and Agentivity at the Twenty-First Regional Meeting*, W. H. Eilfort, P. D. Kroeber, and K. L. Peterson (eds.), 48–63. Chicago: Chicago Linguistic Society.
- 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.
- Himmelman, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36: 161–195.
- Hiramatsu, Kazuko. 2000. Accessing linguistic competence: Evidence from children's and adults' acceptability judgments. PhD diss., University of Connecticut.
- Jackendoff, R. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Jackendoff, R. S. 1983. *Semantics and Cognition*. Cambridge, MA: MIT Press.
- Kay, Paul. 2006. Methodological issues in cross-language color naming. In *Language, Culture, and Society*, Christine Jourdan and Kevin Tuite (eds.), 115–134. Cambridge: Cambridge University Press.
- Kay, Paul, and Brent Berlin. 1997. Science ≠ imperialism. *Behavioral and Brain Sciences* 20: 196–201.
- Kita, Sotaro. 1999. Japanese ENTER/EXIT verbs without motion semantics. *Studies in Language* 23: 307–330.
- Kita, Sotaro. 2006. A grammar of space in Japanese. In *Grammars of Space*, S. C. Levinson and D. P. Wilkins (eds.), 437–474. Cambridge: Cambridge University Press.
- Klein, Wolfgang. 1994. *Time in Language*. London: Routledge.
- Krifka, Manfred. 2011. Varieties of semantic evidence. In *Semantics. An International Handbook of Natural Language Meaning*, vol. I, Claudia Maienborn, Klaus von Stechow, and Paul Portner (eds.), 242–267. Berlin: Mouton de Gruyter.
- Lakoff, George. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic* 2: 458–508.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Levin, Beth. 1993. *English Verb Classes and Alternations*. Chicago: University of Chicago Press.
- Levinson, Stephen C. 2001. Motion verb stimulus, version 2. In *'Manual' for the Field Season 2001*, S. C. Levinson and N. J. Enfield (eds.), 9–11. Nijmegen: Max Planck Institute for Psycholinguistics.
- Lewis, Richard L., Shravan Vasishth, and Julie A. Van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science* 10 (10): 447–454.
- Lucy, John A. 1997. The linguistics of “color.” In *Color Categories in Thought and Language*, C. L. Hardin and Luisa Maffi (eds.), 320–346. Cambridge: Cambridge University Press.
- Magnus, Hugo. 1877. *Die geschichtliche Entwicklung des Farbensinnes* [The Historic Development of the Color Sense]. Leipzig: Viet.

- Magnus, Hugo. 1880. *Untersuchungen ueber den Farbensinn der Naturvoelker* [Investigations on the Color Sense of the Primitive Peoples]. Jena: Fraher.
- Majid, Asifa, James S. Boster, and Melissa Bowerman. 2008. The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition* 109 (2): 235–250.
- Marshall, John C., and Freda Newcornbe. 1973. Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research* 2: 175–199.
- Marslen-Wilson, William, Lorraine Komisarjevsky Tyler, and Rachele Waksler. 1994. Morphology and meaning in the English mental lexicon. *Psychological Review* 215 (1): 3–33.
- Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70: 369–415.
- Meyer, David E., Roger W. Schvaneveldt, and Margaret G. Ruddy. 1974. Functions of graphic and phonemic codes in visual word-recognition. *Memory & Cognition* 2: 309–321.
- Myers, Charles S. 1904. The taste-names of primitive peoples. *British Journal of Psychology* 1: 117–126.
- O'Meara, Carolyn, and Gabriela Pérez Báez. 2011. Spatial frames of reference in Mesoamerican languages. *Language Sciences* 33: 837–852.
- Pederson, Eric, Eve Danziger, David P. Wilkins, Stephen C. Levinson, Sotaro Kita, and Gunter Senft. 1998. Semantic typology and spatial conceptualization. *Language* 74: 557–589.
- Perea, Manuel, and Eva Rosa. 2002a. Does the proportion of associatively related pairs modulate the associative priming effect at very brief stimulus-onset asynchronies? *Acta Psychologica* 110: 103–124.
- Perea, Manuel, and Eva Rosa. 2002b. The effects of associative and semantic priming in the lexical decision task. *Psychological Research* 66: 180–194.
- Piaget, Jean, and Bärbel Inhelder. 1956. *The Child's Conception of Space*. London: Routledge.
- Plaut, David C., and Tim Shallice. 1993. Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology* 10(5): 377–500.
- Rosch, Eleanor. 1978. Principles of Categorization. In *Cognition and Categorization*, Eleanor Rosch and Barbara B. Lloyd (eds.), 27–48. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rosch, Eleanor, and Carolyn B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7: 573–605.
- Saunders, Barbara A. C., and Jaap van Brakel. 1997. Are there nontrivial constraints on colour categorization? *Behavioral and Brain Sciences* 20: 167–228.
- Snyder, William. 2000. An experimental investigation of syntactic satiation effects. *Linguistic Inquiry* 31(3): 575–582.
- Quine, W. V. O. 1960. *Word and Object*. MIT Press.
- Talmy, Leonard. 2000. *Toward a Cognitive Semantics*. Vol. I. Cambridge, MA; London, England: MIT Press.
- Tonhauser, Judith, David Beaver, Craig Roberts, and Mandy Simons. 2013. Toward a taxonomy of projective content. *Language* 89(1): 66–109.
- Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* LXVI: 143–160.
- Wilkins, David P. 1997. Alternative representations of space: Arrernte narratives in sand. In *Proceedings of the CLS Opening Academic Year '97/'98*, Monique Biemans and Joost van der Weijer (eds.), 133–162.
- Wittgenstein, Ludwig. 1921/1922. *Tractatus Logico-Philosophicus*. Trans. C. K. Ogden. London: Routledge and Kegan Paul.