Gathering semantic data[*]

1. Cognitive linguistics and empirical approaches to the study of linguistic meaning

This chapter surveys data collection methods for the study of linguistic meaning. The broader context in which I wish to frame this overview is that of *empirical* approaches to semantic research, i.e., approaches that do not rely on the researcher's first (or "native") language speaker intuitions. This introductory section introduces this perspective and briefly considers its place in the cognitive linguistics enterprise.

How can one know what a linguistic sign or expression means? There are fundamentally two answers to this question. To the extent that the researcher is a competent speaker of (the appropriate dialect and register of) the language to which the expression belongs, she has an understanding of what it means, and she might conceive of her scholarly task as an explication of this understanding. This is a **hermeneutic** approach to the study of meaning. **Empirical** approaches instead take as their point of departure the observation of the communicative behavior of the members of the speech community. Researchers make inferences about the meanings of the observed expressions and test predictions derived from the inferences through further observation, much like psychologists empirically study the mind through observation and hypothesis testing. In practice, both types of approaches have coexisted for as long as researchers have studied linguistic meaning. For example, language documentation and description and the study of child language development generate a myriad of semantic questions that can only be answered empirically. Corpus linguists study the meanings of expressions 'distributionally,' by analyzing the contexts in which the expressions occur, as epitomized by Firth's (1957: 11) famous dictum *You shall know a word by the company it keeps* – this is an empirical approach. However, there has until recently not been an attempt at comprehensively laying out the methodological and epistemological foundations of empirical semantic research (but see Bohnemeyer in press). And there is a widespread misconception, including among linguists, according to which the hermeneutic method is the only method of semantic research. This misconception is likely the result of a confluence of two historic developments: first, linguistics is a social science with roots in the humanities, especially philology and philosophy, where hermeneutic methods dominate; and secondly, Chomsky's (1965, 1966) ideas about introspection of native speaker intuitions as the primary basis of evidence in the study of languages have influenced the training of generations of linguists, including the training of scholars who went on to embrace cognitive linguistics.

Under the influence of rationalists such as Chomsky, linguists, and cognitive scientists more broadly, have been adding new chapters to the debate between rationalism and empiricism that has animated Western thought since the 16th Century. Chomsky (1966) explicitly places his introspection-based research program in the rationalist tradition of what he calls 'Cartesian linguistics'. In this context, the appeal to native speaker intuitions is

---

[*] This chapter is a marginally revised version of a chapter of my forthcoming book *Semantic research: An empirical introduction*. It is reproduced here with permission by the publisher, Cambridge University Press. This book was until recently co-authored by David Wilkins, who in the end decided to withdraw from the project due to his desire to minimize his involvement in academic work past his retirement. I am enormously grateful for his expertise and inspiration. In turn, the textbook chapter incorporates from Section 4 onwards Bohnemeyer (2015), reproduced with permission by Oxford University Press. I have used the textbook manuscript in my Semantics 1 course at the University at Buffalo since 2012 and the text in its present form has benefited very substantially from the comments of many students.

grounded in the postulate of innate knowledge. Even if only a small core part of the sum total of knowledge accessible to (and through) intuition is actually assumed to be innate, it is this assumption of innateness, and of the absence of learnability, that lends the introspection of intuitions primacy over sensory experience as a source of knowledge.

In contrast, cognitive linguistics, with its emphasis on usage-based views of linguistic knowledge and on an embodied understanding of the mind, has in my view a natural affinity with philosophical empiricism. Usage is best studied through observation, as usage manifests itself in socially distributed events in time and space. And the embodied mind is a learning mind, a mind that is attached to the physical world as it receives its primary data through the senses.

The central questions of an empirical epistemology and methodology for semantics are the following:

(1) Which observable properties of communicative behavior can serve as valid data in 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?

The present chapter addresses (2). It discusses the tools – the *methods* – the researcher has at her disposal for collecting data from one or more of sources of evidence. §2 presents a classification of linguistic data and data gathering techniques. §3 discusses the application of corpus methods in semantic research. §4 presents a case study that affords a comparison of the strengths and weaknesses of semantic research based on elicitation versus on spontaneous observation. §5 proposes a classification of elicitation techniques. **Semasiological** approaches, used to elucidate the meanings of given expressions, are discussed in §5; **onomasiological** approaches, which explore the expression of given meanings, in §6. §7 examines the role of hermeneutic methods in empirical research on linguistic meaning. §8 summarizes.

2. Approaches to linguistic data gathering

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 nonlinguistic representation intended as the input of the task. 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 elicitation stimulus) is a speech act directed at the participant(s) by the researcher intended to trigger a set of cognitive computations involving the semantic system. These computations are intended to ultimately result in a response, in such a way that the computations, the representations involved in them, and the speaker's knowledge and practices involved in them can be recovered from the response. 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 nonlinguistic action that solves the task, for example by pointing out a possible referent, demonstrating an action that would instantiate a given description, etc. Figure 1 presents a cartoon version of a rather pedestrian example: the field researcher asks a speaker of Yucatec Maya how to say 'I've got 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 hierarchy 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.

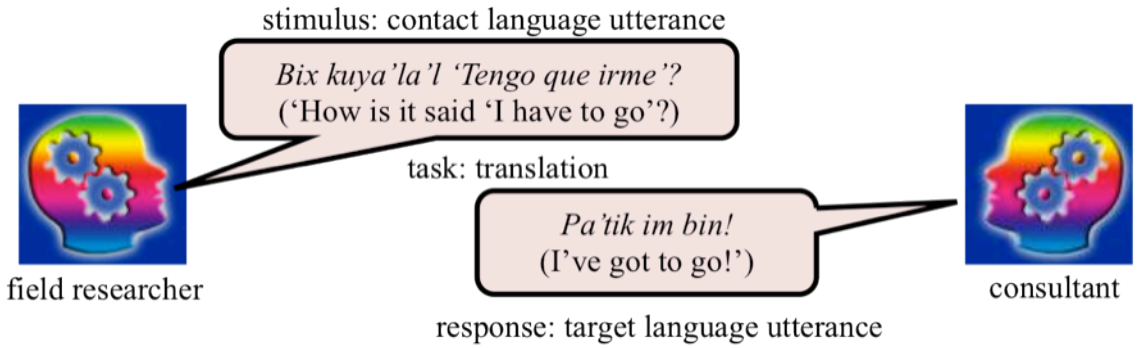

Figure 1. *Components of linguistic data gathering*

Table 1. *The families of data gathering techniques in linguistics*

	Recording of spontaneous speech events	Recording of interviews and 'staged' speech events	Elicitation/experimentation
Linguistic behavior ('response')	+	+	+
Task	-	+	+
Stimulus	-	-	+

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 a cline. 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.

The present section is dedicated to the role of samples of recorded or written spontaneous or staged speech events in semantic research.[1] I first briefly discuss research methods for the study of recorded/written spontaneous and staged speech samples and then offer a comparison of the respective strengths and limitations of semantic research based on recorded/written spontaneous or staged speech samples vs. on elicitation.

3. Corpus methods in semantics

In the language sciences, a **corpus** may be defined as a set of speech samples that are typically, though not necessarily, of a common origin or source and that have been prepared for further analysis in some unified fashion. Preparation usually, though not necessarily, includes **transcription** and optionally various forms of annotation. Annotation that identifies and categorizes elements of linguistic code – turns, sentences, clauses, phrases, words, or morphemes, optionally along with their semantic or phonological properties – is known as **tagging**, especially when it is inserted automatically by some program or script and/or when it is provided as mark-up for generic research purposes without a specific study goal in mind. This contrasts (somewhat informally; this terminological dichotomy is not accorded a great deal of significance as far as I am aware) with **coding**, the identification and categorization of expressions with certain (phonological, morphosyntactic, semantic, or pragmatic) properties for a specific research purpose.

By the above definition, **corpus data** is not restricted to corpora that consist of recorded or written speech samples. Elicited data may be prepared and presented in a corpus as well. And the methods for studying corpus data are the same regardless of whether the speech samples a corpus consists of were elicited or recorded/written. Nevertheless, the discussion focuses in the following on recorded/written corpus data.

What can a semanticist do with a corpus? The principal type of information a corpus provides is **distributional information**. That is to say, for any given expression – be it a word or word form, a particular morpheme, or a construction – the corpus allows the researcher *in principle* to assess the following:

- In which types of discourse does the expression occur? In particular, in which **genres** of text (e.g., cooking recipes; Congressional hearings; sports broadcasts; marriage counseling sessions; etc.) and in which **registers** (formal; technical; colloquial; etc.) does it occur?
- What are the properties of the linguistic and nonlinguistic context in which the expression occurs? For example, does the expression occur in (response to) a particular type of speech act? Does the expression occur with a referent that has been mentioned before?
- What other expressions does the expression co-occur with (in particular syntactic configurations)?

The relative ease of obtaining answers to these questions depends on the tagging that is available and the extent to which the researcher has to annotate the corpus specifically to obtain the answer. The principal type of data a corpus study can provide in answer to these questions is **frequency data**: the frequency with which an expression occurs in the corpus in a context of particular specifications.

---

[1] By 'written' speech samples, I primarily have in mind speech that was not written in response to a researcher's prompt – newspaper articles, personal letters, novels, and so forth. Of course, elicitation may likewise result in written samples.

One of the most influential corpus studies carried out in linguistics attempted to elucidate the semantic factors underlying the English **dative alternation**, i.e., the choice between double-object datives (*Sally gave Floyd the book*) and prepositional datives (*Sally gave the book to Floyd*). Based on the three-million word Switchboard Corpus of recorded telephone conversations (Godfrey et al 1992), Bresnan et al (2007) and Bresnan & Nikitina (2009), building on precursors such as Thompson (1990), were able to show that the choice of one construction over the other was independently influenced by each of the following factors:

- A previously mentioned referent was preferred to be ordered before a newly introduced one;
- A pronoun was preferred to be placed before a lexically headed phrase;
- An animate referent was preferred to be mentioned before an inanimate referent;
- A definite noun phrase was preferred to be ordered before an indefinite one;
- A shorter expression was preferred to be ordered before a longer one.

Logistic regression models[2] that took into account all of these variables (without ranking) were able to correctly predict 92% of the uses of the two constructions in the corpus.

The study by Bresnan and colleagues illustrates that the power of corpus data, particularly in combination with sophisticated statistical analyses, lies in giving the researcher unparalleled access to **variation** in the use of a particular expression and the factors that govern this variation, be they lexical, discourse-pragmatic, stylistic, or sociolinguistic factors. Moreover, this representation of variation is based on observations of the actual use of the expression, as opposed to speaker's judgments about when they use the expression, which do not always match the actual use and in fact can be a surprisingly poor representation of the latter (as I illustrate in §5).

Bresnan et al. (2007) discuss some common objections against the use of corpus data. One important limitation of spontaneous speech samples is that they generally fail to provide direct **negative evidence** regarding the applicability of an expression to a given referent in a given context (at least failing a serendipitous **metapragmatic** comment by one of the speakers in the corpus). One can argue that if an expression is never used in reference to a certain kind of situation in a corpus the size of the Switchboard Corpus, this is a strong approximation of negative evidence. However, corpora of this size do not presently exist for the vast majority of languages.[3] Moreover, as I illustrate in §4 and §5, the non-occurrence of

---

[2] Regression analysis is a family of statistical (or machine learning) algorithms that are designed to find the closest approximation to a given dataset that can be represented as a function of a given set of variables. Such an approximation is called a 'regression model'. The model estimates a coefficient for each variable, the degree of independence/correlation among the variables, and some measure of the probability that the actual coefficient falls within a certain confidence interval of the estimated coefficient. Various subtypes of regression analysis are distinguished in terms of the algebraic type of function/model they aim to 'fit' to the data. The most common subtypes are linear and logistic regression analysis.

[3] There is also currently no list or database of corpora of the languages of the world. The vast majority of these corpora are simply the data collections of individual researchers. However, there is now a sustained and coordinated push toward creating archives that make such datasets available to the community of researchers and the community of speakers over the internet, with endangered languages being given priority status in this efforts. These include the Archive of the Indigenous Languages of Latin America at the University of Texas (AILLA; http://www.ailla.utexas.org); the Endangered Languages Archive at the School of

a particular expression in reference to a particular kind of situation may be due to pragmatic rather than semantic factors. An appropriate context in which these pragmatic factors are absent may prove that the expression is in fact semantically compatible with the situation. But that type of context may be outside the makeup of a particular corpus. Bohnemeyer (2012), to be discussed in the following section, provides an illustration of the importance of negative evidence for semantic research.

4. Conversational vs. elicited data: a case study

This section further examines the no-negative-evidence problem of research based on unelicited data, but also makes the case that elicitation that wants to do justice to the language under study must be informed by unelicited data. For a practical illustration, consider Bohnemeyer (2012), which offers a comparison of a widely known study of the Yucatec demonstrative system by Hanks (1990, 2005) with results obtained in fieldwork applying the Demonstrative Questionnaire developed by David Wilkins at the Max Planck Institute for Psycholinguistics.[4] Hanks' 'practice' approach is based on recordings of spontaneous interactions in culturally typical settings such as the household, cornfield, religious ceremonies, etc. He developed coding schemas for the participants when deictically referring to places and objects in these settings. These coding schemas showed the spatial layout of the settings, and Hanks assigned numbers to certain prominent locations and objects to code the interactions he recorded in the settings. Based on these data, he carried out a detailed analysis of the *use* of the demonstrative forms in the interactions he had observed. However, he did not attempt to go beyond usage and venture into an analysis of the underlying semantics of the forms.

4.1. Hanks (1990, 2005)

Hanks avoids the terms 'proximal' and 'distal', arguing that these are "obscured in standard approaches to deixis which take as their touchstone 'real' space rather than social interaction" (488). Instead, he uses the labels 'immediate' and 'non-immediate', which Bohnemeyer (2012) adopts.

The Yucatec system of spatial deixis is fairly complex and includes some typologically highly unusual traits. A brief sketch will be useful. Among determiners, the immediate- vs. non-immediate opposition is the only opposition there is. But in the adverbial system, the immediate-non-immediate opposition between *te'la'* 'here' and *te'lo'* 'there' semantically intersects with an 'inclusive-exclusive' opposition between *waye'* 'here' and *tolo'* 'there'. In other words, there are two 'here's and two 'there's in Yucatec. Hanks calls the 'inclusive-exclusive' opposition between *waye'* and *tolo'* 'egocentric'. This distinction presupposes some kind of perimeter around the speaker, such that *waye'* refers to the inside of that perimeter and *tolo'* to its outside. The perimeter can be defined by the boundaries of for example the house, the field, the village, or the state where the conversation takes place. The addressee is normally inside the perimeter as well. *Tolo'* is used in indiscriminate reference to things that are "out there" in the relevant respect. Table 2 summarizes Hanks' analysis of the space-deictic determiners and adverbs of Yucatec.

Oriental and African Studies (ELAR; http://elar.soas.ac.uk/); and The Language Archive at the Max Planck Institute for Psycholinguistics (TLA; https://tla.mpi.nl/). The Aboriginal and Torres Strait Islander Data Archive (ATSIDA; http://www.atsida.edu.au), a part of the Australian Data Archive (ADA), has a somewhat broader mission not restricted to linguistic data.

[4] There has been a considerable amount of research on spatial deixis in cognitive linguistics. Cf. in particular Talmy (2017) and references therein.

Table 2. *The semantics of the space-deictic determiners and adverbs of Yucatec according to Hanks (1990)*

Meaning	Inclusive		Exclusive
	Immediate	Non-Immediate	
Form class			
Adverbs	*way ...=e'* 'here'		*tol ...=o'* 'there'
	te'l ...=a' 'here'	*te'l ...=o'* 'there'	
Determiners	*lel=a'* 'this one' *le ...=a'* 'this'	*lel=o'* 'that one' *le ...=o'* 'that'	

Waye' 'here' and *tolo'* '(out) there' cannot normally be contrasted in reference to places that speaker and addressee have visual access to, as such places would be within the perimeter and hence entirely inside the domain of *waye'*. Similarly, if there are multiple possible referents for *waye'*, they are concentric and thus cannot easily be distinguished gesturally. Therefore, both terms can be used without accompanying gestures, and the only gestures that do accompany them are gestures that do not point to specific places.[5]

In contrast to the 'egocentric' 'inclusive-exclusive' distinction, the 'immediate-non-immediate' opposition between *te'l ...=a'* 'here' and *te'l ...=o'* 'there' and the determiners *lela'/le ...=a'* 'this' and *lelo'/le ...=o'* constitutes what Hanks calls a 'sociocentric' system. He observes that these forms are used *contrastively* with respect to speaker and addressee, respectively: immediate forms are used for reference to objects or places closer to the speaker than to the addressee, while non-immediate forms are used in reference to objects or places closer to the addressee.

Hanks notes that the usage patterns his analysis ascribes to the immediate and non-immediate forms differ "in two details: (i) the relative remoteness of the (…) possible referents, and (ii) the foregrounding of the addressee rather than the speaker. The second feature is motivated by the fairly consistent association between the 'there' of *te'lo'* and the addressee's location" (Hanks 1990: 437). Consider some of the examples that Hanks quotes in support of this analysis. These are examples in which speaker and addressee are in relatively close proximity, such as (4)–(5) in which a child is chided by an adult while both are in the same room and in the second case even less than two meters apart. Yet the speaker picks the non-immediate form to refer to the child's location:

(4) *Mak* *a=chi'* ***te'l=o'***, *páal!*
Close(B3SG) A2=mouth there=D2 child
'Shut up **over there**, kid!' (Hanks 1990: 438)

(5) *Ts'a'* *le=ba'l* ***te'l=o'***!
Give/put(B3SG) DET=thing there=D2
'Put that thing down **there**!' (Hanks 1990: 438)

[5] The 'egocentric' terms play only a marginal role in responses to the Demonstrative Questionnaire; therefore, they are not discussed further. *Way ...=e'* 'here' did not occur at all, and *tol ...=o'* 'out there' only occurred once in a consultant's first response. Interestingly, the two scenes one would predict to be most likely to trigger *tol ...=o'* based on Hanks' 'perimeter' analysis, 20 and 21, failed to elicit *tol ...=o'*.

This raises the question whether the non-immediate terms, i.e. the determiner *lelo'* / *le …=o'* and the place adverb combination *te'lo' / te'l …=o'*, are actually addressee-based; i.e., whether they encode proximity to the addressee, rather than distance from the speaker. An example of a language with an addressee-based demonstrative is Japanese. Japanese has a demonstrative *ko* for referents close to the speaker, a demonstrative *so* for referents close to the addressee, and a demonstrative *a* for referents that are in the proximity of neither the speaker nor the addressee.[6] Addressee-based terms like Japanese *so* are found somewhat regularly in three-term demonstrative systems; they compete with other types of three-term systems that distinguish three degrees of distance from the speaker or two degrees plus one distance-neutral term, as in the case of Turkish.

Applying an addressee-based analysis to the immediate-non-immediate contrast in Yucatec straight away runs into the problem that the latter is a binary contrast. So one would have *lela'* / *le …=a'* or the adverb combination *te'la'* / *te'l …=a'* for entities and places close to the speaker, *lelo' / le …=o'* or the adverb combination *te'lo'* / *te'l …=o'* for entities and places close to the addressee, and then the question arises as to what to use for entities and places that are neither in the speaker's nor in the addressee's zone of proximity. For this reason, a two-term demonstrative or deictic adverb system is not very likely to include an addressee-based term; and indeed, the typological surveys of Anderson and Keenan (1985) and Diessel (1999) do not include a single example of such a system – only three-or-more-term systems may include addressee-based terms. However, a two-term system with one speaker-based and one addressee-based term is by no means *impossible*. One conceivable realization of such a system might be found in a language in which demonstratives or deictic adverbs are simply *not used* in reference to objects that are neither close to the speaker nor to the addressee. The hypothetical language would employ other means to this end, such as explicit locative descriptions. But this is very clearly not the case in the dialect of Yucatec discussed here.[7]

Hanks (1990: 490) in fact observes that the *=o'* forms are used in reference to entities and places in both the addressee's zone *and* the 'common field'. It is not completely clear to us how this 'common field' is to be construed (see Enfield 2003 for a possibly similar analysis). At any rate, the 'common field' would presumably cover a significant part of the space outside both speaker's and addressee's 'zones' (i.e., areas of proximity).

Leaving aside the issue of how entities outside the 'common field' would be referred to, the main question that arises is how to reconcile the 'foregrounding' of the addressee by the non-immediate forms with the fact that they are also used for reference to objects and places in the common field outside the addressee's zone. Hanks (2005) suggests that this foregrounding is a pragmatic rather than semantic effect:

---

[6] This is a simplified account based on unpublished research by Sotaro Kita. See Kita & Walsh Dickey (1998: 66) and Senft & Smits (2000: 69) for summaries.

[7] As mentioned above, the 'egocentric' adverb *tol …=o'* 'out there' is according to Hanks used for vague reference to places outside some perimeter around the deictic center. In a hypothetical two-term system with forms for the speaker's and the addressee's zones, this would indeed be a solution to the problem of referring to objects and places that are in neither zone. However, it would be a solution only for those special circumstances in which *tol …=o'* is used (i.e., there *is* a salient perimeter around the deictic center, and the reference object/place *is* situated outside it. As likewise mentioned above, *tol …=o'* plays only a marginal role in the responses to the Demonstrative Questionnaire.

> The rule of thumb is therefore simply, in pragmatically contrastive contexts such as greetings and scoldings, to treat [the speaker's] field as *a'* and [the addressee's] field as *o'*. **When I state this association as a rule of thumb I mean to underscore that it is not part of the semantics of Yucatec deixis, since it is easy to find examples in which the association is cancelled**. It is, however, part of the routine handling of types of exchange that happen throughout any ordinary day. (Hanks 2005: 206; emphasis *JB*)

Hanks' use of the term 'cancellation' suggests that his "rule of thumb" is a Gricean stereotype implicature, i.e., that the addressee's zone of proximity is in many instances the stereotypical search domain of the non-immediate forms. This, however, implies that those "easy to find" situations in which the non-immediate forms are used in exophoric reference to entities or places outside the addressee's zone of proximity are somewhat less typical. It is one of the strengths of elicitation approaches such as the one presented in the following that they permit the realization and testing of reference in such atypical situations. This puts the researcher in a position to distinguish between semantic and pragmatic meaning components. In order to determine the role of the addressee's location in the use of the non-immediate forms, their use needs to be examined in contexts in which the relative locations of speaker, addressee, and reference object are systematically varied. Controlling these variables is one of the main goals of the Demonstrative Questionnaire.

4.2. Bohnemeyer (2012)

The Demonstrative Questionnaire (Wilkins 1999) gives instructions for the enactment of 25 scenarios, specifying for each scenario the relative locations of the participants and the object to be referred to, but also the referent's status in discourse and the object's status with respect to the interlocutors' focus of attention. This is an instance of the method of controlled elicitation with non-verbal stimuli, which plays a crucial role in **semantic typology**, the crosslinguistic study of semantic categorization (cf. Evans 2010; Moore et al. 2015).

The variables controlled in the Demonstrative Questionnaire were of course determined on the basis of prior research. Hanks' (1990) influential study of demonstrative use in spontaneous interactions in Yucatec was among the sources that were considered in the design of the questionnaire. A major goal of Hanks (1990, 2005) is to show that the meaning and use of demonstratives are primarily governed by interactional variables rather than by purely spatial properties such as in particular measurable distance. In the design of the Demonstrative Questionnaire, both spatial and interactional variables are controlled for. The descriptions specify for each scene a setting (e.g., inside a walled-off space; on a ballgame field); a spatial configuration of speaker, addressee, and reference object, and optionally a bystander, within that setting; the kind of reference object at issue (one of the speaker's teeth, a bug, a radio, book, or ball); and a number of additional properties such as whether joint attention between speaker and addressee is on the referent at the moment of utterance or is rather directed to it by the speaker in the course of the utterance, whether the object has been mentioned before in the course of the conversation, and whether the object is owned (in whichever sense) by one of the interlocutors. The spatial configurations vary the distances between speaker, addressee, reference object, and bystander and the visibility and accessibility of the object from the vantage point of speaker and hearer. Distance from speaker and/or addressee is varied in terms of a seven-point scale, according to which the object is a body part vs. in contact with the body vs. within arm's reach vs. within easy access a few steps away vs. tens of meters away vs. more than a hundred meters away vs. several

kilometers away.[8] The descriptions are realized as verbal instructions to the researcher supported by diagrams; Figures 2–4 below show examples of these diagrams.

During the enactment, a native speaker consultant is meant to assume the role of speaker and another or the researcher that of the addressee. The researcher describes the scene for the speaker, records the utterance the speaker considers most appropriate in each scenario and/or the range of utterances the speaker considers acceptable, and optionally asks follow-up questions to clarify properties of the elicited utterances and/or test the influence of additional variables.

The Yucatec questionnaire data were collected in August 1999 with five adult native speakers, four men and one woman, aged between 25 and 52. All spoke Yucatec as their first and dominant language, but were fluent to varying degrees in Spanish as well. The 25 questionnaire scenes were enacted with the consultants as speakers and me as addressee. The enactments were conducted at the appropriate scale except for the far-distant scenes 13–18 and 24–25, which were enacted at a reduced scale. In order to judge the significance of the data (given the small number of consultants), it will be worth pointing out that the five consultants generally showed a high degree of convergence in their responses. For example, in their first choices between an immediate and a non-immediate form (regardless of whether they also considered a form of the complementary set applicable, and whether they volunteered that other form or merely agreed to its applicability), all five consultants agreed with respect to 15 of the 25 scenes; and only three scenes elicited a two-to-three split in this regard. Moreover, in two of the three scenes that elicited the largest amount of variation, scenes 2 and 4, in fact all consultants agreed that both immediate and non-immediate forms would be applicable, depending on the proximity between the speaker's pointing gesture and the reference object. This suggests that the data do in fact permit viable generalizations about the knowledge of Yucatec native speakers regarding the use of demonstrative forms in exophoric spatial reference.

The Yucatec questionnaire study failed to find any evidence suggesting that the relative location of the addressee with respect to the speaker or the reference object has any direct impact on the selection of forms for exophoric reference. It is not even the case that non-immediate forms are applied more readily and/or consistently in reference to objects and places close to the addressee than in reference to objects or places distant from both speaker and addressee. The results of the questionnaire study thus do not support the hypothesis that the addressee's zone of proximity constitutes a focal area within the extension of the non-immediate forms.

To determine the impact of the addressee's location on the choice of deictic forms, responses to scenes that only differ in the addressee's location need to be compared, such as scenes 13 and 16, depicted in Figure 2. In both scenes, the speaker and the reference object are on opposite ends of a football field, but the addressee is very close to the speaker and far away from the object in one case and very close to the object and far away from the speaker in the other case. All five consultants unanimously use non-immediate forms under both conditions, regardless of the location of the addressee. A typical response is (6):

(6) *Le=ràadyo=**o'*** *(yàan te'l=**o'**),* *hach* *ma'+lóob.*
DEF=radio=**D2** EXIST(B3SG)there=**D2** really NEG+bad(B3SG)
'That radio (that is over there) is really nice'

[8] Scenes in which the object is equidistant from speaker and addressee vary distance according to the last five of these seven points.

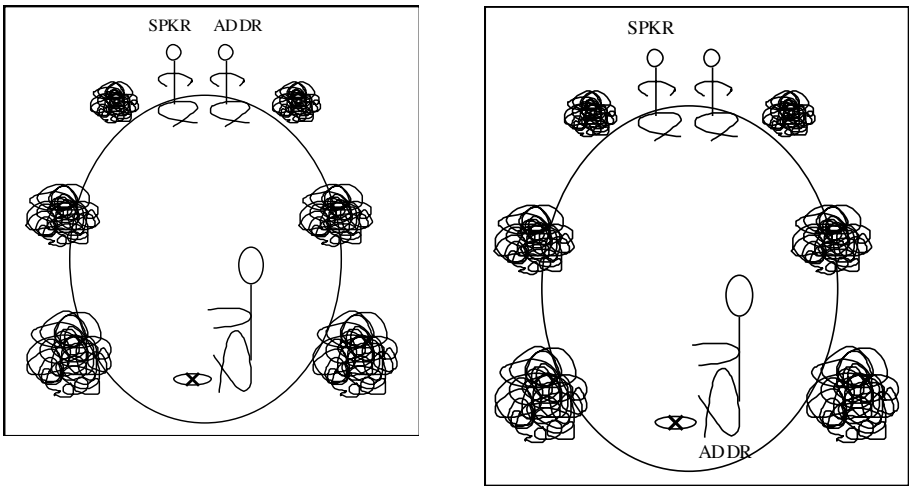

Figure 2. *Demonstrative Scenes 13 (left) and 16*

A non-immediate form is optionally augmented by the deictic locative adverb *te'l*. The choice of whether or not the more complex form is used depends on the attention variable. The consultants just as readily used the non-immediate forms in reference to an object distant from both speaker and addressee in 13 than they did in reference to an object close to the addressee in 16; there is thus no evidence suggesting that places and objects in the addressee's proximity play any special role in the reference of the non-immediate forms. A similar point can be made with respect to scenes 9 and 12, depicted in Figure 3. In 9, the reference object is close to the addressee and out of the speaker's reach. In 12, the object is equidistant from speaker and addressee and out of either's reach.

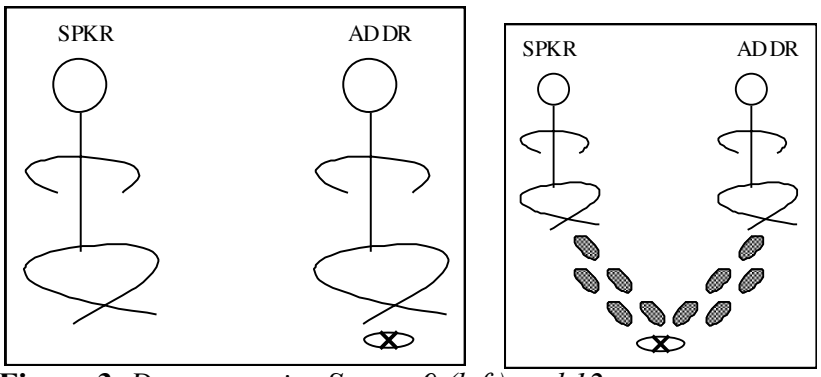

Figure 3. *Demonstrative Scenes 9 (left) and 12*

Given Hanks' observations of a privileged association between the non-immediate forms and the addressee's zone, it may have been expected that the non-immediate forms are more readily applied in 9 than in 12. And on the hypothesis that reference to the addressee's zone is the semantic prototype of the non-immediate forms, this is in fact clearly predicted. But if anything, the opposite is the case: all five consultants prefer non-immediate forms in 12, but only four out of five do so in 9. Again, both simple non-immediate forms and augmented constructions were used, depending on whether the addressee's attention was assumed to be on the object prior to the utterance. The augmented form in this case is formed with the presentative adverb *he'l*, not with the locative adverb *te'l*. A typical example is (7).

(7) *A=ti'a'l* *le=lìibro* *(he'l)=o'?*
A2=property(B3SG) DEF=book PRSV=D2
'Is that book (there) yours?'

Having failed to find a direct addressee bias in the use of the non-immediate forms, the hypothesis that the non-immediate forms are semantically specified for exophoric reference to places and objects outside the speaker's proximity needs to be considered. A glance at the overall distribution of immediate vs. non-immediate choices across the 25 Demonstrative Scenes as presented in the appendix shows that this cannot be correct. Figure 4 summarizes the consultants' responses to eight of the scenes. Solid lines represent a clear preference; dotted lines cases in which the particular forms were considered acceptable but not preferred.

A single scene in which the use of non-immediate forms is excluded – reference to the speaker's own body in scene 1 – contrasts with no less than 12 scenes that exempt the use of immediate forms. A response to scene 1 is reproduced in (8). All five speakers rejected the use of non-immediate forms in this context.

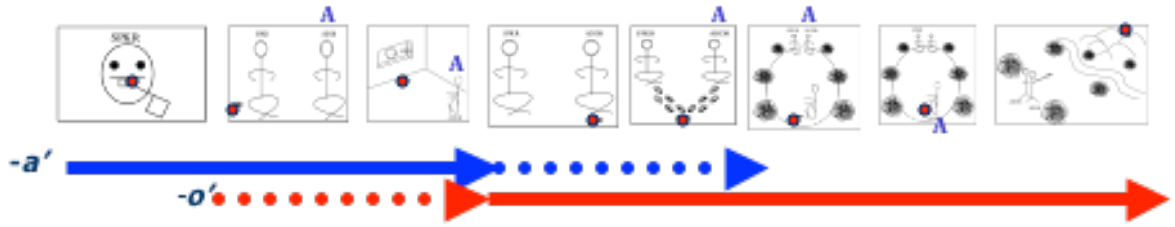

Figure 4. *Preferred (solid) and accepted (dotted) responses across eight of the scene. The blue arrow represents the extension of the immediate forms, the red arrow that of the non-immediate ones. The blue letter A marks the position of the addressee in the scenes and the red dot that of the referent.*

(8) *Tèen=e' mu'n bèey-tal*
me=D3 NEG.PROSP:A3 thus-INCH.INC
in=meyah, tuméen túun ki'nam
A1SG=work CAUSE PROG:A3 hurt
in=koh he'l=a'
A1SG=tooth PRSV=D1
'Me, I can't work, because this tooth here of mine is hurting'

Non-immediate forms infringe on immediate territory all the way up to that first scene, while the immediate forms are clearly confined to the speaker's zone of proximity. This distribution suggests that the immediate forms are semantically 'marked' vis-à-vis the non-immediate ones, i.e., that their use is more restricted because their meaning is narrower. Only the immediate forms are semantically specified for exophoric reference to a particular region, namely, the speaker's zone of proximity. But what, then, is the common denominator in the uses of the non-immediate forms?

Perhaps the non-immediate forms express 'neutral deixis', i.e. exophoric reference without restriction to a particular region of space. However, the non-immediate forms are also used for anaphoric reference and definiteness marking, as illustrated in (9) and (10):

(9) *Ba'x k'iin k-uy=úuch-ul* ***lel=o'***?
what sun IMPF-A3=happen-INC DEF=D2
'What day does **that** usually happen?'

(10) *Káa=h-òok*

káa=PRV-enter(B3SG)
le*=*x-ch'úup ***chak*** ***u=nòok'*=*o'***, *(...)*
DEF=F-female red(B3SG) A3=garment=D2
'(And then) **the woman** dressed in red entered, (...)'

The non-immediate particle is obligatory in the contexts illustrated in (9)-(10), although other anaphoric and definite expressions – in particular, 3rd-person pronouns and proper nouns – do not trigger it. In light of these **endophoric** uses of the non-immediate forms, the neutral-deixis hypothesis can only be maintained under an additional assumption of polysemy. Therefore, in the absence of further evidence, Occam's Razor appears to favor an analysis of the non-immediate forms as generic indexicals which do not semantically distinguish between exophoric and endophoric reference.

Given that the non-immediate forms are semantically neutral regarding the immediate-non-immediate contrast, why are they dispreferred for reference to objects/places in the speaker's proximity? The semantic analysis just outlined cannot explain this, so the answer has to be sought in the pragmatics of the system. Bohnemeyer (2012) argues for a traditional Gricean **preemption** analysis: a generalized conversational implicature based on Grice's (1975) First Maxim of Quantity ("Make your contribution as informative as is required") or Levinson's (2000) equivalent 'Q-heuristic' ("What isn't said, isn't"). This mechanism yields an inference to the non-applicability of the marked term wherever the marked term is not chosen. In the case of the spatial deictics of Yucatec, preemption generates a default interpretation of the non-immediate forms according to which they do not refer to objects/places in the speaker's proximity, based on the reasoning that if the speaker were in fact referring to his or her region of proximity, why would (s)he not use an immediate form, given that the immediate forms are positively specified for this reference? This mechanism is invoked in the analyses of demonstrative systems proposed by Enfield (2003) and Levinson (2006).[9]

However, the preemption analysis does not immediately account for the exemption of the non-immediate forms from reference to the speaker's own body, as in Scene 1. This exemption in fact extends to objects pointed at or touched by the speaker at close range, as in Scenes 2 and 4 of the Questionnaire. Consultants consider the use of non-immediate forms in these contexts decidedly odd. It seems conceivable that similar phenomena may be encountered in bona-fide cases of preemption as well. Perhaps the flip side of Grice's "Make your contribution as informative as is required" is that speakers who use an informationally weaker term where evidence for the validity of the additional conditions of the stronger term is clearly available are felt to not make their contribution informative enough. For example, to say 'Steve ate some of the cookies' famously implicates but does not entail that Steve did not eat all of the cookies. But to say this holding the empty cookie jar (Steve still chewing) might well be interpreted, although not strictly false, as inaccurate in some contexts.

4.3. Comparison of the two approaches

Let us now summarize and evaluate the comparison of the two approaches. Hanks' 'Practice' approach is based on the observation of deictic usage in spontaneously occurring interactions. As shown, this approach falls somewhat short of clearly establishing category boundaries or cutoff points. In the case at hand, Hanks observes a frequent association between the use of non-immediate forms and the addressee's location. Hanks (2005) clarifies that this

[9] Fillmore (1997) suggests a similar analysis to explain why *Tuesday* is inferred to not mean 'today' in case it is uttered on a Tuesday.

association is not a *semantic* property of the non-immediate forms. The questionnaire study has clearly confirmed this: the non-immediate forms are readily applied to places and entities outside the addressee's proximity and in fact even in the speaker's proximity, right up to the boundaries of the speaker's body. However, Hanks' analysis still suggests that the addressee's zone of proximity plays a privileged role in the *use* of the non-immediate forms. Perhaps the addressee's zone of proximity is a focal area in the extension of the non-immediate forms and their interpretation is therefore biased toward this focal point by stereotype implicatures. But the questionnaire study has also failed to produce any evidence in support of such typicality effects: the non-immediate forms were just as readily and consistently used in reference to entities and places outside the addressee's zone of proximity as they were used with referents that were near the addressee. This suggests that the apparent addressee bias Hanks observed may in fact be nothing more than a statistical correlation: the non-immediate forms were used most frequently with referents close to the addressee simply because Hanks' database of observed interactions is biased toward referents that are close to speaker, addressee, or both. This conclusion should be understood as tentative and preliminary: it should be checked, on the one hand, against a quantitative analysis of Hanks' database and, on the other, against psycholinguistic studies of the production and comprehension of Yucatec demonstrative forms. If there is an addressee bias, it should manifest itself for example in word association tests and in faster processing of non-immediate forms when used in reference to the addressee's region of proximity compared to when used in reference to other places.

The Demonstrative Questionnaire approach is based on controlled elicitation of usage under artificial conditions. This method offers the following closely related principal advantages over the observation of spontaneous data:

- Elicitation can generate evidence of how speakers and hearers behave in situations that occur less commonly or even marginally in spontaneous interactions. Such evidence can provide important clues about the underlying categories, representations, and procedural knowledge speakers and hearers rely on. To put it in more general and abstract terms, any scientific analysis seeks to describe and explain the dependencies between the variables that affect the phenomena at issue. Such an analysis will be incomplete unless it covers all possible variable-value combinations, including combinations that are difficult to study except under artificial conditions. Reference to entities and places that are neither in the speaker's nor in the addressee's zone of proximity in the present study is arguably a case in point.

- Realization of all possible variable-value combinations in a grid-like design may help avoid misinterpretations of statistical correlations as causal relations. This is exemplified in the present study by the association between the non-immediate forms and the addressee's location in Hanks' analysis, which the questionnaire study has identified as a possible result of a statistical bias in Hanks' database.

- A grid design may also help uncover systematic relations between variables that otherwise elude the researcher due to their complex nature. A case in point is the systematic correlation between attention direction and the use of augmented forms the questionnaire study has shown.

- Finally, the elicitation of native speaker *judgments* may produce *negative* evidence of particular variable-value combinations not only not occurring spontaneously, but being excluded from occurrence due to ungrammaticality, semantic anomaly, or

pragmatic infelicity, or a combination thereof. It may also show certain combination to be acceptable but non-idiomatic. Conversely, showing that certain infrequent and/or atypical combinations are nevertheless possible may help identify pragmatic implicatures. In the questionnaire study, this latter principle was applied twice. It failed to support a stereotype implicature analysis of the apparent addressee bias in the use of the non-immediate forms in Hanks' data. But it did produce evidence of a scalar implicature or preemption effect pragmatically excluding the non-immediate forms from the speaker's zone of proximity.

On the downside, one drawback of elicitation is that it is by itself blind, as it were. A questionnaire design such as Wilkins' needs to be informed by specific research questions and will only provide answers to the questions implemented. As mentioned above, Hanks' meticulous study of demonstrative use in spontaneous interactions was in fact an important source in the design of the Demonstrative Questionnaire. Beyond this, Hanks (1990, 2005) makes a number of intriguing observations that the questionnaire study failed to replicate, simply because the relevant variable was not implemented in the questionnaire. For instance, many of Hanks' examples refer to motion events rather than stative locations. Hanks (1990: 432–433) notes that speakers consistently use immediate forms in reference to motion goals they are en route to and non-immediate forms in reference to motion sources that they have already left. The Demonstrative Questionnaire has no way of detecting this phenomenon, because motion is not coded in the demonstrative scenes.

The second principled drawback of elicitation is that it only determines what native speakers do under simulated conditions. This immediately raises questions of **validity**. One aspect where the Demonstrative Questionnaire proves artificial in a way that may well limit the validity of any study conducted with it has to do with the role of joint attention and attention direction in demonstrative usage. The questionnaire study has produced evidence suggesting that attention direction is grammaticalized in the Yucatec systems of spatial deixis. While the simple immediate and non-immediate forms are used when a joint focus of attention on the reference object or place has been established prior to the reference act, complex forms augmented with the presentative adverb *he'l* or the place adverb *te'l* are used to direct the addressee's attention to the reference object or place. The choice between the *te'l* forms and the *he'l* forms depends not on the physical accessibility of the reference object or place, as with the choice between immediate and non-immediate forms, but on identifiability of the reference object/place in the visual field. However, attention is coded in the questionnaire in instructions to ask the consultant to *imagine* that the researcher as the addressee is, say, not aware of the reference object, that (s)he may not have noticed it, etc. In essence, this means asking the consultant to imagine that (s)he is not thinking about something! Obviously, this is methodologically unsatisfactory. But controlling the focus of attention under experimental conditions is an extremely difficult task. In the absence of a technique for doing this (see Enfield and Bohnemeyer 2001 for a possible solution), the observation of natural interactions may be our best bet in the study of the role of attention direction in spatial deixis.

The upshot of this comparison of the two approaches to the study of spatial demonstratives, observation of spontaneous interactions and controlled elicitation, seems clear enough: to ensure optimal results, the two are best pursued in tandem. Moreover, the observation of spontaneously occurring speech must naturally take the lead role in this combination. Elicitation is essentially a cleanup job that helps to sort out and make sense of the results of spontaneous observation.

5. Elicitation: from meaning to utterance

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

5.1. Elicitation and experimentation

As already indicated above, linguistic **elicitation** can be defined as the collection of responses to linguistic or nonlinguistic stimuli designed to study the respondents' linguistic competence and/or their practices of language use. 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. Elicitation and experimentation are not mutually exclusive. They can be thought of as cluster concepts that have distinct prototypes, but overlap in their extensions. Another possible view on the distinction is that 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 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.

5.2. A classification of elicitation techniques

A classification of the elicitation techniques at the linguist's disposal can be achieved by cross-tabulating 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 3. This presupposes that the identification of stimulus ("input") and response ("output") types alone is sufficient to define the task types. I am unaware of a compelling theoretical reason why this should be so, but it seems to work out quite nicely.

Table 3 does not list psycholinguistic experimental techniques that primarily yield processing data, e.g., chronometric or attentional measures, since these do not produce linguistic data and can only indirectly contribute to studies of linguistic meaning.

The classification in Table 3 distinguishes four stimulus types. The stimulus is either an utterance - in which case it may be a target language utterance or a contact language utterance – or the content of some linguistic or nonlinguistic 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 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.

Table 3. *A classification of elicitation techniques by stimulus and response type*

Response **Stimulus**	Target language utterance	Contact language utterance	Metalinguistic utterance: judgment	Metalinguistic utterance: description	Nonlinguistic 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 nonlinguistic representation	**Type IV –** description				

The response in turn can be a target language utterance, a contact language utterance, a metalinguistic utterance, or some nonlinguistic 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, felicity, goodness of fit as a representation of a particular state of affairs, etc., on a scale, whereas the latter paraphrase its meaning or describe a scenario or a setting, etc., in which it might be used.

As Table 3 shows, the only stimuli that can be used to elicit all of the five response types are utterances in the target language. At the same time, utterances in the target language also constitute the one valid response type that may be elicited using *any* stimulus type .So utterances in the language of study necessarily play a role in any kind of elicitation, whether it is as stimulus, response, or both. This makes sense given that target language utterances are the primary data of semantic research. 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 3 covers not only the methods for semantic elicitation, but 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 that aims to identify how a given meaning is expressed, **completion** and **association tasks**, **translation tasks**, **contextualized production tasks**, and **description tasks** are suitable. Conversely, for studies aiming to elucidate what a given expression means, eliciting **judgments** of (non-) contradiction, felicity, etc., **explications** by paraphrase or scenario, and **demonstration** and **act-out tasks** will be the methods of choice.

In the following, illustrations of the seven types are provided using examples from my 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 3. The remainder of this section is dedicated to techniques that target expressions of a given meaning, whereas §6 deals with the inverse direction of inquiry.

5.3. 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 & Keyser 1987); cf. Bohnemeyer 2007; Majid et al 2008). My objective in this study was to determine which verbs impose narrow selectional restrictions 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 (11) and the **conative alternation** in the case of the *cut*-class (12) (Fillmore 1967; Levin 1993; Levin & Rappaport-Hovav 1995: 89-119; *inter alia*). As (13) indicates, the classes of verbs that participate in these alternations do not overlap.

(11) a. *Floyd broke the vase*
b. *The vase broke*
(12) a. *Sally cut the bread*
b. *Sally cut at the bread*
(13) a. **The bread cut*
b. **Floyd broke at the bread*

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 (14) and a typical- instrument prompt of the format in (15) (i administered the task in Yucatec, using Yucatec prompts):

(14) '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?'

(15) '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 4 and 5 list the responses for *hat* 'tear' and *xot* 'cut':

Table 4. *Responses for* hat *'tear'*

Responses to theme prompt (14)	Responses to instrument prompt (15)
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 5. *Responses for* xot *'cut'*

Responses to theme prompt (14)	Responses to instrument prompt (15)
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, or pieces torn off an aluminum can

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 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 & Landar (1963) and Clark (1970). A very interesting recent application can be found in Evans & Wilkins (2000).

5.4. Type II and III: Story vignettes

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'[10] 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

[10] Textual deixis is the use of spatial deictics (demonstratives and space-deictic adverbs) in reference to portions of the linguistic context of the utterance in which they appear, as in *This is a short example embedded in a long sentence*, where *this* refers to the sentence in italics that contains it. Cf. Bohnemeyer (2015), Levinson (2004), and references therein.

stimuli and finite verb forms are replaced with non-finite ones (set in capital letters to flag them) wherever possible.[11] Instead, the intended perspective is controlled by a context that defines a reference time or 'topic time' (Klein 1994) for the translation stimulus. The topic time is the time about which an utterance makes a statement or asks a question, etc. This is illustrated in (16). The context precedes the translation stimulus and is set in brackets.

(16) 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 (16) 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.[12] 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 (17) is a Yucatec translation of the stimulus in (16). To ensure that the speaker takes the context fully into account during the translation, I asked the speaker to translate the context as well.

(17)
Q: Ba'x a=tukul-ik k-u=beet-ik
what(B3SG) A2=think-INC(B3SG) IMPF-A3=do-INC(B3SG)
'What do you think he will be (lit. is) doing'
a=suku'n chéen k'uch-uk-o'n?
A2=elder.brother SR.IRR arrive-SUBJ-B1PL
'your big brother when we arrive?'
A: Chéen k'uch-uk-o'n wal=e',
SR.IRR arrive-SUBJ-B1PL UNCERT=D3
'When we arrive, I guess'
ts'íib-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.

5.5. 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 & Kay (1969) (with much earlier precursors such as Magnus (1877, 1880) in research on the linguistic categorization of color

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

[12] A 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 the **etic grid problem**: the set of semantic or notional categories which the study is designed to test – the study's **etic grid** – biases the possible observations of semantic categories in the target language. Cf. Moore et al. (2015) for discussion.

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 us illustrate the role of the task first. Consider Figure 5, 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 & Imhelder (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.

However, when I ran BowPed with Yucatec speakers, I had an experience that has been reported by several other researchers: in response to the question 'Where is the cup?', a speaker would look at the researcher with mild puzzlement and point to the picture: 'Uh, right here?' What this response suggests 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,** i.e., a more elaborate context within which to respond, a context that put a certain interpretation on the *Where*-question:

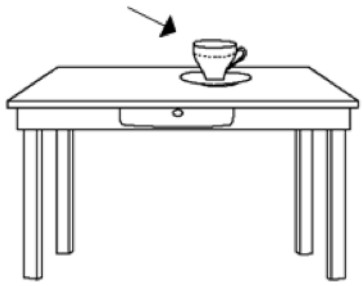

Figure 5. *Item #1 of the "Topological Relations Picture Series" aka BowPed (©Eric Pederson; reproduced with permission)*

(18) "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?"

He repeated this frame with every new picture until he 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 & Wilkes-Gibbs 1990). They involve at least two speakers per trial: one describes the content of a stimulus and the other re-matches the description to a set of non-linguistic stimuli. There are numerous possible realizations of this, including picture to picture, picture to toy model, etc. Figure 6 illustrates the setup of a picture-to-picture matching task.

The "Ball & Chair" referential communication task (Bohnemeyer 2011) was developed to replace and improve upon a similar task, "Man & 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. Spatial reference frames are cognitive "coordinate" systems – technically, sets of axes (cf. Levinson 1996, 2003; Bohnemeyer & O'Meara 2012; inter alia) – that serve to define regions and directions in space. Consider the scene in Figure 7 below: from the perspective of the observer, and in this case also from the perspective (as it were) of the chair, the ball is 'left of' the chair. Were the chair facing the observer, the ball would be on its right. This is an example of an 'intrinsic' frame, whereas the observer perspective involves a 'relative' frame. In a suitable 'absolute' or 'geocentric' frame, one might say, for example, that the ball is 'west of' or 'downhill from' the chair. Geocentric descriptions of small-scale space sound odd to Westerners, but are not uncommon in many non-Western cultures.

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 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 (19) reproduces a description of Figure 7 in full.

Figure 6. *Setup of the Ball & Chair picture matching task*

Figure 7. *Ball & Chair picture 2.5*

(19) a. Estée, u séegere-e-e … chan fòotoa’,
esté u=séegir le=chan fòoto=a’,
HESIT A3=follow DEF=DIM photo=D1
‘Uh, this next-uh-little photo,’

b. u frèente e sìiyao’, tu tohile don Jorgeo’,
u=frèente le=sìiya=o’ tu=tohil le=don Jorge=o’
A3=front DEF=chair=D2 PREP:A3=straight:REL DEF=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 (towards) 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’éel bòola=i’
PREP=EXIST(B3SG) one-CL.IN ball=D4
‘there is a ball.’

g. Ta xts’íi - ta xts’íihk’abil [unintel.], ti’ yàan ump’éel bòolai’,
ta=x-ts’íik+k’ab-il=e’ ti’=yàan hun-p’éel bòola=i’
PREP:A2=F-left+hand-REL=TOP PREP=EXIST(B3SG) one-CL.IN ball=D4
‘On your le – on your left [unintelligible], there is a ball,’

h. kàasi tu tohil u yòok yàan ti’.
kàasi tu=tohil uy=òok yàan ti’=i’
almost PREP:A3=straight:REL A3=leg/foot EXIST(B3SG) PREP(B3)=D4
‘it’s almost in the line of its leg with respect to it.’

Line g is a correction of line e. The individual propositions of this description can be analyzed under the assumption that they are true of the described stimulus item, i.e., the picture in Figure 7. *X-ts’íik* ‘left’ in line g could be ambiguous with respect to Figure 7, permitting both a relative interpretation projected from the body of the addressee and an intrinsic one projected from the chair itself as reference entity or ‘ground’ of the locative description. However, the morphologically bound 2nd-person possessor pronoun rules the second interpretation out, making it clear that line g involves a relative frame of reference.

Referential communication tasks offer advantages over more traditional, more controlled forms of elicitation of permitting the observation of the linguistic strategies interlocutors use to coordinate on a set of stimuli. Studying these strategies is otherwise mostly restricted to spontaneous observation, where the researcher has no control over the content of the interaction. However, referential communication tasks share with all forms of elicitation the imposition of a context on the interaction that is not very representative of everyday interactions in the life of the community. In referential communication tasks in particular, this lack of ‘ecological validity’ manifests itself above all in blocking the participants from effective use of co-speech gesture. I take up ecological validity, the extent to which an observation in the social and behavioral sciences is representative of behavior that occur outside the study context, in the next subsection.

5.6. Ecological validity

In discussions of the topic of using referential communication designs in field research, concerns about *ecological validity* are regularly voiced. Ecological validity is one of a family of criteria used to evaluate scientific study designs. The most prominent members of this family are *internal validity* and *external validity*. **Internal validity** assesses the extent to which a study design is suitable for addressing the underlying research question and the extent to which a study supports the proposed conclusions. **External validity** measures the extent to which the findings of a study generalize to the larger population from which the study sample is drawn and to situations markedly different from the conditions of the study. **Ecological validity** is the degree to which a study permits conclusions about the everyday behavior of the participants outside the conditions of the study. The status of ecological validity in science and the relation between external and ecological validity are controversial. A proper discussion of the issues goes well beyond the scope of this chapter.

There are two aspects to the ecological validity problem in referential communication designs: the artificiality or unfamiliarity of the stimulus, and that of the task. The former problem pertains to any research with non-linguistic stimuli that are alien to the culture of the speech community, whereas the latter is more or less a unique property of referential communication tasks. Let us consider the more specific issue first.

Participants unfamiliar with advanced information technology may not be very accustomed to using speech in contexts where gaze and gesture cannot serve to disambiguate referents. More importantly, few people are accustomed to communicating detailed spatial information in such contexts. Except for visually impaired speakers and highly technical genres of communication, the conveyance of rich small-scale spatial information naturally relies heavily on gaze and gesture. What this means is that responses to a task such as Ball & Chair can tell us something about the cognitive and communicative resources that the

members of a given community tap into when faced with an unfamiliar task of certain specifications, but they do not permit a direct assessment of the actual practices of language use in the community. To make this more concrete: the description in (19) involves three spatial reference frames, a frame anchored to the researcher's body standing near the camera that recorded the description of Figure 7 ('in the line of that don Jorge') and two frames anchored to the body of the addressee or a generic observer, one that does not involve projection of the body's axes onto the chair ('turned toward your front') and one that does ('on your left', i.e., on the observer's left of the chair). What this shows is that this particular speaker is capable of using these kinds of frames in reference to small-scale space. Furthermore, if the description results in a successful match, this suggests that the addressee is capable of using the same frames in the comprehension of the speaker's descriptions. Next, by analyzing the total set of descriptions by a particular speaker, we can assess that speaker's preferences among the strategies available to them for solving this artificial task. By comparing 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 & 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 – e.g., in the case of spatial reference, route descriptions – and non-elicited, i.e., staged discourses (first-hand 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.

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, etc. 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 8, 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 birds-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.[13] (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 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.

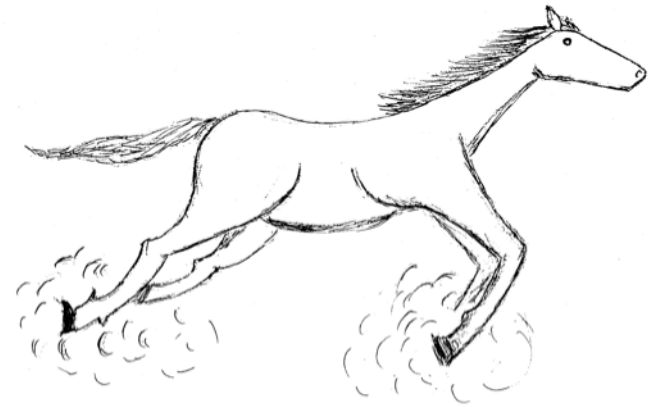

Figure 8. *Galloping horse or dead horse? (Wilkins 1997: 157; ©David P. Wilkins)*

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

[13] Sand drawing on the camp ground are of course exclusively viewed from above – hence the naturalness of the birds-eye perspective.

speaker's interpretation of it. This is captured by the Golden Rule of Elicitation, to be further discussed below.

Another often-commented-on limitation of nonlinguistic 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.

6. Elicitation: from utterance to meaning

The discussion of elicitation methods continues within techniques that help elucidate the meanings of given utterances.

6.1. 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,[14] 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 state 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 observed 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, etc. 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. Such solutions tend to rely on workarounds that will not necessarily work for all languages and communities, but that are sufficiently nontechnical to ensure availability across a broad range of languages and cultures.

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. For example, to test whether (20a) entails (20b), the researcher might construct and test (20c).

---

[14] **Idiomaticity** can be defined as the extent to which an otherwise compositional complex expression has fixed, non-compositional, conventional meaning components. Consider for example the contrast between (i) 'What time is it?' and (ii) 'How late is it?'. Example (ii) is the literal English translation of the German equivalent of (i). To a German speaker, a literal German translation of (i) makes about as little sense as a literal English translation of (ii) does to an English speaker. Both might argue that their native approach to asking this question is the "logical" one, i.e., the compositional one. From the perspective of a logician or semanticist, both expressions are partially compositional and partially have *idiomatic* meanings, i.e., are factually interpreted as idioms.

(20) a. *The painting is above the door*
b. *The door is below the painting*
c. *The painting is above the door, but the door is not below the painting*

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 almost 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. This is illustrated by (20c): if the propositions expressed by the two clauses can be true simultaneously in the same situation, then (20a) does not entail (20b). However, in actual fat, (20c) appears to be a contradiction, suggesting that (5.3.1a) does in fact entail (20b).

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, etc. 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 7 have correct and incorrect solutions and that the correct ones employ cardinal direction terms. The origin of this belief is at present unclear.

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 structures (e.g., combination of words or constructions) that appear initially anomalous sometimes seems to improve with time as hearers are exposed to instances of the same structure over and over 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.[15]

The ability to distinguish between different types of anomaly – as induced by syntactic vs. semantic vs. pragmatic clashes, etc. – depends on a consultant's declarative, metalinguistic understanding of linguistic phenomena and thus grows with the consultant's experience and training. 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. A valid strategy is 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 are listed in (21):

(21)
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 (21a-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, (21e-h) can be used to test whether a given description is accurate and pragmatically appropriate in reference to a particular non-verbal stimulus (21e-f) or a verbally described scenario (21g-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 non-verbal 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

[15] Frequency appears to influence speech processing heavily and might thus at least indirectly be involved in satiation effects.

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 non-verbally represented scenarios, beginning with the former type. Our example for this comes from testing Yucatec verb phrases for telicity. **Telicity** is the property of certain event descriptions (telic descriptions) to require realization of a final part of the kind of eventuality they lexicalize in order for the description to be applicable. Consider the examples in (22) below. An activity of drawing a circle doesn't become an instance of the action described by *draw a circle* until the circle is complete. In contrast, as soon as somebody has started pushing a cart, they have already done something that can be described as instance of *push a cart*, an atelic description. Unlike in English, telicity has no syntactic reflexes in Yucatec (Bohnemeyer 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 Bohnemeyer & 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 term). As illustrated in (22), **activity** (i.e., atelic dynamic) descriptions in the progressive entail their simple-tense counterparts (22a), whereas the same does not hold for **accomplishment** (telic durative) descriptions in the progressive (22b):

(22) 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 negotiated with five native speaker consultants 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.

(23)
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'o'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 responded negatively to (23), since *kàay* 'sing', the antipassive stem of the transitive root *k'ay* 'sing', is normally interpreted as 'sing a song' (cf. Bohnemeyer 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, Bohnemeyer (2010) 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) videos. 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 (24) acceptable in reference to the clip whose first and last frame are depicted in Figure 9, in which a plank slides under a ball and cylinder:

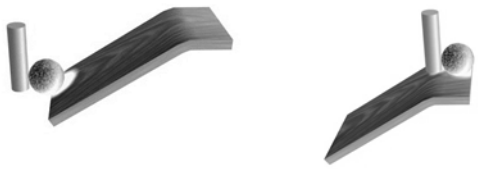

Figure 9. *First and last frame of "FIGURE_GROUND 14" (Levinson 2001; ©Stephen C. Levinson; reproduced with permission)*

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

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 (24), 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. The discussion of this study continues in the next subsection.

6.2. 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

(24) above by adding a scenario that would turn it into an acceptable description of the scene depicted in Figure 9. One response to this procedure is shown in (25):

(25) Le=chan tàabla=o' h=péek-nah-ih,
DEF=DIM plank=D2 PRV=move-CMP-B3SG

káa=h-na'k le=chan kanìika
CON=PRV-ascend(B3SG) DEF=DIM marble

y=éetel che' te'l y=óokol=o'.
A3=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 (24), which clashes with the scenario shown in Figure 9, 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. However, in terms of the semantic content they contribute to utterances, Yucatec verbs such as *na'k* 'ascend' in (24) and (25) seem to lexicalize just location change of the figure vis-à-vis the ground, rather than motion along a path. Whether the location change comes about by the figure moving, the ground moving, or figure or ground coming into or going out of existence in configuration with the other (as in teleportation scenarios) is not specified as part of the semantics of these verbs.

"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.

6.3. Type VII: From target language utterance to nonlinguistic representation

The final type of elicitation task has speakers produce a nonlinguistic representation of the meaning of a target language expression. *Demonstration* seems to us 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 play-acting 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. Distinctions that enter the conceptualization of dispositions include posture and support/suspension (e.g., 'sit', 'stand', 'lie', 'kneel', 'lean', 'hang', 'droop', 'dangle', 'be mounted on top of something'); blockage of motion (e.g., 'be stuck to something', 'be stuck between two things'); orientation in the gravitational field (e.g., 'lie face up', 'lie face down', 'lie on side', 'be tilted at an angle'); and configurations of parts of an object with respect to each other (e.g., 'be scattered', 'be spread out', 'be in a pile', 'be

lined up in a row', 'be bulging', 'be bent', 'be twisted', 'be coiled up'). Mayan languages have hundreds of roots of this kind, and the majority of these select for inanimate figures. For this reason, I prefer 'dispositional' over 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 & 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 Tseltal 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 10 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.

Figure 10. *Suspension dispositions described by (clockwise from top left)* choh*,* ch'uy*,* lech*, and* t'oy

7. 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 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 should in fact in one respect also include an element of hermeneutic analysis in its toolkit in order to achieve its goal. Hermeneutic methods aim to explicate the researcher's understanding of texts (and other representations and semiotic practices) in a manner that is consistent with all aspects of the text and what is known about it.

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 (24) as a description of the animation in Figure 9 did not apparently intend this judgment to be understood to the effect that (24) 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 9, with the result shown in (25).

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.[16]

8. 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.[17]

References

Belloro, Valeria, Jürgen Bohnemeyer, Dedre Gentner, & Kathleen Braun. 2008. Thinking-for-speaking: evidencia a partir de la codificación de disposiciones espaciales en español

[16] A similar view is stated in Matthewson 2004.

[17] Of 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.

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 & Paul Kay. 1969. Basic Color Terms. Berkeley, CA: University of California Press.
Berlin, Brent. 1968. Tzeltal Numeral Classifiers: A Study in Ethnographic Semantics. Janua Linguarum (All Series), Bloomington, IN; 70. The Hague: Mouton de Gruyter.
Bohnemeyer, Jürgen & Mary D. Swift. 2004. Event realization and default aspect. *Linguistics and Philosophy 27* (3): 263-296.
Bohnemeyer, Jürgen & Penelope Brown. 2007. Standing divided: Dispositional verbs and locative predications in two Mayan languages. *Linguistics 45* (5-6): 1105-1151.
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 Yukatek Maya. Doctoral dissertation, Tilburg University.
Bohnemeyer, Jürgen. 2000. Event order in language and cognition. In H. de Hoop & T. van der Wouden (eds.), *Linguistics in the Netherlands 17*. Amsterdam: John Benjamins. 1–16.
Bohnemeyer, Jürgen. 2002. The grammar of time reference in Yukatek Maya. Munich: LINCOM.
Bohnemeyer, Jürgen. 2007. Morpholexical transparency and the argument structure of verbs of cutting and breaking. *Cognitive Linguistics 18* (2): 153-177.
Bohnemeyer, Jürgen. 2010. The language-specificity of Conceptual Structure: Path, Fictive Motion, and time relations. In Barbara Malt & Philip Wolff (eds.), *Words and the world*. Oxford: Oxford University Press. 111-137.
Bohnemeyer, Jürgen. 2011. Spatial frames of reference in Yucatec: Referential promiscuity and task-specificity. *Language Sciences 33*: 892-914.
Bohnemeyer, Jürgen. 2012. Yucatec demonstratives in interaction: spontaneous vs. elicited data. In Andrea C. Schalley (ed.), *Practical theories and empirical practice*. Amsterdam etc.: John Benjamins. 99-128.
Bohnemeyer, Jürgen. 2015. A practical epistemology for semantic elicitation in the field and elsewhere. In R. Bochnak & L. Matthewson (eds.), *Methodologies in semantic fieldwork*. Oxford: Oxford University Press. 13-46.
Bohnemeyer, Jürgen. In press. *Semantic research: An empirical introduction*. Cambridge: Cambridge University Press (2022).
Bowerman, Melissa & Eric Pedersen. Ms. Cross-linguistic perspectives on topological spatial relationships. Manuscript, Max Planck Institute for Psycholinguistics.
Bresnan, J. & T. Nikitina. 2009. The Gradience of the Dative Alternation. In *Reality Exploration and Discovery: Pattern Interaction in Language and Life*, edited by Linda Uyechi and Lian Hee Wee, 161--184. Stanford: CSLI Publications; Pdf.
Bresnan, J., A. Cueni, T. Nikitina, & H. Baayen. 2007. Predicting the Dative Alternation. In *Cognitive Foundations of Interpretation*, ed. by G. Boume, I. Kraemer, and J. Zwarts. Amsterdam: Royal Netherlands Academy of Science, pp. 69--94. 33 pages.
Chamberlain, Alexander Francis. 1903. Primitive taste words. *The American Journal of Psychology 14*: 146–153.
Chomsky, N. A. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
Chomsky, N. A. 1966. *Cartesian linguistics: A chapter in the history of rationalist thought.* Cambridge: Cambridge University Press.

Clark, Herbert H. & Deanna Wilkes-Gibbs. 1990. Referring as a collaborative process. In Philip Cohen, Jerry Morgan, & Martha Pollack (eds.), *Intentions in communication*. Cambridge, MA: MIT Press. 463-493.
Clark, Herbert H. 1970. Word associations and linguistic theory. In John Lyons (ed.), *New horizons in linguistics*. Baltimore, MD: Penguin. 271-286.
Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.
Diessel, Holger. 1999. *Demonstratives: form, function, and grammaticalization.* Amsterdam etc.: Benjamins.
Dowty, David R. 1979. *Word meaning and Montague Grammar: The semantics of verbs and times in Generative Semantics and Montague's PTQ*. Dordrecht: Reidel.
Enfield, Nick J. 2003. Demonstratives in space and interaction: Data from Lao speakers and implications for semantic analysis. *Language* 79(1): 82–117.
Ervin, Susan M. & Herbert Landar. 1963. Navajo word association. *American Journal of Psychology 76*: 49-57.
Evans, N. (2010). Semantic typology. In J. J. Song (Ed.), *The Oxford handbook of linguistic typology*. Oxford: Oxford University Press. 504-533.
Evans, Nicolas & David P. Wilkins. 2000. In the mind's ear: the semantic extensions of perception verbs in Australian languages. *Language 76* (3): 546-592.
Fillmore, Charles J. 1997. *Lectures on deixis*. Stanford, CA: CSLI Publications.
Francom, Jerrid. 2009. Experimental Syntax: Exploring the effect of repeated exposure to anomalous syntactic structure—evidence from rating and reading tasks. PhD thesis, University of Arizona.
Frege, Gottlob. 1980 [1884]. *The Foundations of Arithmetic*. Trans. J. L. Austin. Second Revised Edition. Evanston, Illinois: Northwestern University Press.
Firth, John Rupert. 1957. A synopsis of linguistic theory 1930-1955. In J. R. Firth (ed.), *Studies in linguistic analysis*. Oxford: Blackwell. 1-32.
Godfrey, J., E. Holliman and J. McDaniel. (1992). SWITCHBOARD: Telephone speech corpus for research and development. *Proceedings of ICASSP-92*, 517–520.
Goodall, Grant. (2005). On the syntax and processing of wh-questions in Spanish. In Schmeiser, B., Chand, V., Kelleher, A., & Rodriguez, A. (eds.), *WCCFL 23 Proceedings*. Somerville, MA: Cascadilla Press. 101–114.
Grice, Herbert Paul. 1975. Logic and Conversation. In P. Cole & J. L. Morgan, eds., *Speech acts*, 41-58. New York, NY: Academic Press.
Guerssel, Mohamed, Kenneth Hale, Mary Laughren, Beth Levin, & Josie White Eagle. 1985. A cross-linguistic study of transitivity alternations. In W. H. Eilfort, P. D. Kroeber, & K. L. Peterson (eds.), *Papers from the parasession on causatives and agentivity at the twenty-first regional meeting*. Chicago, IL: Chicago Linguistic Society, 48-63.
Hale, Kenneth L. & Samuel J. Keyser. 1987. A view from the middle. *Lexicon Project Working Papers, 10*. Cambridge, MA: Center for Cognitive Science, MIT.
Hanks, William F. 1990. *Referential practice: language and lived space among the Maya*. Chicago etc.: University of Chicago Press.
Hanks, William F. 2005. Explorations in the deictic field. *Current Anthropology* 46(2): 191–220.
Himmelmann, 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 thesis, 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 & Brent Berlin. 1997. Science ≠ imperialism. *Behavioral and Brain Sciences* 20: 196-201.
Kay, Paul. 2006. Methodological issues in cross-language color naming. In Christine Jourdan & Kevin Tuite (eds.), *Language, culture, and society*. Cambridge: Cambridge University Press. 115-134.
Kita, Sotaro & Laura Walsh Dickey, eds. 1998. *Annual report 1998*. Nijmegen: Max Planck Institute for Psycholinguistics.
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 S. C. Levinson & D. P. Wilkins (eds.), *Grammars of space*. Cambridge: Cambridge University Press. 437-474.
Klein, Wolfgang. 1994. *Time in language*. London: Routledge.
Levin, Beth. 1993. *English Verb Classes and Alternations*. Chicago, IL: University of Chicago Press.
Levinson, Stephen C. 2000. *Presumptive Meanings: The Theory of Generalised Conversational Implicatures*. Cambridge, MA: MIT Press.
Levinson, Stephen C. 2001. Motion verb stimulus, version 2. In S. C. Levinson & N. J. Enfield (eds.), *'Manual' for the field season 2001*. Nijmegen: Max Planck Institute for Psycholinguistics. 9-11.
Levinson, Stephen C. 2006. The language of space in Yélî Dnye. In Stephen C. Levinson & David P. Wilkins, eds., *Grammars of space,* 157–205. 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, & Melissa Bowerman. 2008. The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition 109* (2): 235–250.
Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International Journal of American Linguistics 70*: 369-415.
Moore, R. E., K. T. Donelson, A. Eggleston, & J. Bohnemeyer. (2015). Semantic typology: New approaches to crosslinguistic variation in language and cognition. *Linguistics Vanguard* 1(1): 189-200.
Myers, Charles S. 1904. The taste-names of primitive peoples. *British Journal of Psychology 1*: 117–126.
O'Meara, Carolyn & 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, & Gunter Senft. 1998. Semantic typology and spatial conceptualization. *Language 74*: 557-589.
Piaget, Jean & Bärbel Imhelder. 1956. *The child's conception of space*. London: Routledge.
Senft, Gunter. & Roel Smits, eds. 2000. *Annual report 2000*. Nijmegen: Max Planck Institute for Psycholinguistics.
Senft, Gunter. 2004. Aspects of spatial deixis in Kilivila. In Gunter Senft (ed.), *Deixis and demonstratives in Oceanic languages*, 59-80. Canberra: Pacific Linguistics.
Snyder, William. 2000. An experimental investigation of syntactic satiation effects. *Linguistic Inquiry* 31 (3): 575-582.
Talmy, Leonard. 2000. *Toward a cognitive semantics. Volume I*. Cambridge, MA: MIT Press.
Talmy, Leonard. 2017. *The targeting system of language.* Cambridge, MA: MIT Press.

Thompson, S. (1990). Information flow and dative shift in English discourse. In J.A. Edmondson, F. Crawford and P. Mühlhäusler (eds), *Development and diversity: Language variation across space and time*. Dallas, TX: Summer Institute of Linguistics. 239–253.

Tonhauser, Judith, David Beaver, Craige Roberts, & 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 Monique Biemans & Joost van der Weijer (eds.), *Proceedings of the CLS opening academic year '97/'98*. 133-162.

Wilkins, David P. 1999. 1999 Demonstrative Questionnaire: "THIS" and "THAT" in comparative perspective. In David P. Wilkins, ed., *"Manual" for the 1999 Field Season*, 1-24. Nijmegen: MPI for Psycholinguistics.

Wittgenstein, Ludwig. 1921/1922. *Tractatus Logico-Philosophicus*. Trans. C. K. Ogden. London: Routledge & Kegan Paul.