**Experiments in "Language and Thought"**

Herbert H. Clark

In this handout I describe how to carry out and write up an experiment for Psychology 131, *Language and Thought*. I first describe what the paper should look like and then I offer a potpourri of methods for carrying out such an experiment.

## Experimental Write-up

The write-up of your experiment should take the form of a short journal article, in journal article style, although (obviously) without the detail, rigor, or size of such an article. (For examples, see *Psychological Science*, *Journal of Memory and Language*, *Memory and Cognition*, or *Journal of Personality and Social Psychology*.) The paper should be at least 8 and no longer than 20 pages long. It should include four sections:

1. ***Introduction. Why did you do this study****?* Give the background for the experiment. It is best to begin with at least one article, book, or chapter related to the notions you are testing; some references will be suggested in the experiments described later. In particular, describe and justify the theoretical issue you are trying to resolve. Ordinarily, you will give a statement of the hypotheses you want to test. It is these hypotheses that motivate your methods.

   Here is a point that I cannot emphasize too strongly: Every experiment requires a **comparison** of at least two conditions—for example, an experimental condition vs. a control condition. If you don't have a comparison, you don't have an experiment. Be sure you make it clear what your hypothesis predicts about that comparison.

2. ***Methods. What did you do?*** Describe the methods by which you carried out your experiment. Describe the materials used, the number of participants tested, what you asked them to do, etc. Some methods don't rely on participants, but you still need to specify what you did. Someone reading your "Methods" section should be able to replicate exactly what you did.

3. ***Results. What did you find?*** Describe your results—highlighting your comparisons—by presenting the data you collected **in summary form**. Include averages, percent correct, average number of errors—whatever summary numbers are appropriate for your data. Ordinarily, you would present these numbers in tables or graphs, clearly labeled for where the numbers came from and what they represent. (MS Word, which most of you use, has excellent graphing capabilities; so does Excel.) Explain in enough detail that a reader can figure out precisely what you found. You may want to include your raw data in an appendix to the paper.

4. ***Discussion. What does it all mean?*** Explain how the experimental results do or do not resolve the issue you raised in the Introduction—how they support or disconfirm your hypotheses. You might also explain what the results imply for language and thought more generally, what further experiments are needed to decide the issue, what possible artifacts you see in your methods or results, and other such questions. By the end, we readers should know what you think we should take away from the experiment and with what confidence.

Most of you won't have the equipment or statistical expertise for a genuine journal article, so I will be looking for the thought and effort that went into the paper. Good marks will go to experiments that are carefully done, write-ups that are comprehensible and interesting, and discussions that deal the experimental implications, even when the results are inconclusive. The best marks will go to those experiments that are inspired in some way—e.g., in the hypothesis or idea that went into them.

One more comment about writing. Only one half of science is carrying out the research; the other half is making the results public. Scientific writing is too often incomprehensible or boring, and it doesn't have to be. The best writing is clear, readable, persuasive. Even though this isn't an English course, take the writing seriously. I will. You don't need to be flashy or poetic. Just be clear, readable, and precise. Make every word count. (See my essay on how to write psychology papers: "Everyone can write better (and you are no exception).")

In what follows I describe a range of techniques for doing experiments on conversation, production, and understanding. I have included only a few techniques in each area. There are other techniques, of course, and you should seek them out if none of these suit your purpose. I have included a technique here only if: (1) it works; (2) it is doable with simple equipment (e.g., nothing more complicated than an audio or video recorder); and (3) it has theoretical interest. If you are thinking about other techniques, you might check them out with us first.

## Referential Communication

Investigating conversation usually requires an audio or video recorder and a good idea of what you are looking for. (There are audio recorders on most Macs and PCs, and video recorders on every smart phone. There are excellent programs for processing audio such as *Audacity (audacity.sourceforge.net)*.) Studying conversation can get difficult quickly if you have to transcribe much of it—that takes a *lot* of time, let me tell you—so find a project where you need to transcribe selectively, if at all. Ordinarily, you will have to do more than record random conversations. Here are a few suggestions:

In the so-called referential communication task, one participant (called the

director) is given a set of, say, 12 Tangram figures (complicated difficult to describe block figures) on 3X5 cards, or 12 picture post cards, or other figures, and another participant (the matcher) is given an identical set. The two partners can hear but not see each other. The director's cards are placed in a fixed order, the matcher's in a random order, and the two partners have to talk about the cards one at a time in order to get the matcher's in the same order s the director's. You can then rearrange the director's cards and have them do the whole procedure again. On the first trial (arranging all 12 cards), the two partners will take a long time, but on later trials, they will get faster because their references will get shorter. It is easy to measure, from the recordings, the time taken per figure, or per trial. It is more work to transcribe what they said, but not too difficult to transcribe selective parts.

This task allows you to look at many hypotheses. Here are some hypotheses that have *already* been looked at (most of them by my students and me) and others that haven't been looked at.

(1) If an expert about, say, New York is talking to a novice (a non-New Yorker), and if they have to arrange postcards of New York landmarks, they should quickly discover their discrepant expertise and go down to the lowest level of expertise, the novice's, using few if any proper names (e.g., "Empire State Building") by themselves. (See Isaacs and Clark reference.)

(2) Two partners will speed up over six trials. However, if you replace one of the two partners for six more trials, the next two partners will slow back down as if they were starting over (from their own common ground). (Wilkes-Gibbs & Clark)

(3) Overhearers who are not allowed to talk to the director and matcher make more errors in arranging the cards than the matcher. (Schober & Clark)

(4) Directors take *much* longer when the matcher isn't allowed to talk, giving feedback to establish common ground as they go along. They also take longer under other restrictions. (Krauss & Weinheimer)

(5) Two partners should have much more difficulty when only some of their figures match—that is, they cannot take as common ground that their two sets of figures are identical. What should change?

(6) When two partners can see each other (but not their cards), they are likely to use iconic (or depictive) gestures along with their language—or *instead* of their language. If you have a video camera, time, and patience, you could examine the gestures people use in the Tangram task. Here are some possible questions: (a) Do two partners gesture more when they *can* see each other than when they cannot? Are the gestures when they can see each other the same as when they cannot see each other? (b) Do two partners gesture more when they are just starting than later on when they have short names for the Tangram figures? (c) If you follow

the gestures used for one or two Tangram figures over trials, how do they change over trials? (I suspect they get simpler, but how and why?)

Many other hypotheses are possible. Some depend on the figures being arranged—e.g., how much the partners know about them, how discriminable the figures are, what they consist of, etc. Others depend on whether the partners can see each other and gesture, whether they are forced to use single words, etc. For the basic techniques, see (all available on SULAIR or my website):

Clark, H.H., & Wilkes-Gibbs, D. (1986) Referring as a collaborative process. *Cognition*, 22, 1-39.

Isaacs, H.H., & Clark, H.H. (1987) References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116, 26-37.

Krauss, R.M., & Weinheimer, S. (1964) Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, 1, 112-114.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21, 211-232.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Cognition*, 31, 183-194.

**Hesitations, Pauses, and Repairs**

Experiments on hesitations, pauses, and repairs need an audio recorder too. The basic technique is to record a person's speech, transcribe parts of it, and then have yourself and perhaps one other person judge the hesitations, pauses, and repairs in the speech, marking the transcription where each of these occurs. Again, as illustration, here are some techniques that have already been tried and others that haven't (that I know of).

(1) One can look at spontaneous conversations under two or more contrasting conditions, and examine the number and types of repairs used. For example, two people may yield a different pattern of interruptions, repairs, and the like when talking about a topic one person is expert in than when talking about a topic both are experts in, or when one person has authority over the other. A server at a counter or restaurant may speak differently, and with a different pattern of repairs and hesitations, to a person with a foreign accent than to a person with a local accent.

(2) One can compare people trying to describe simple vs. complicated figures. One might expect difficulties in ordering what to describe in the more

complicated case, with different sorts of repairs. (See the Levelt reference.)

(3) Think of people talking will driving. Their talk surely slows down, or becomes disfluent, when they run into serious traffic problems. You can simulate this by having people talk to each other while one or both participants are required periodically to do an extra intellectual task—like take the next turn in a computer game or play the next card in a computer card game. Disfluencies should be greater the more demanding the secondary task, and people should make more errors on the secondary task.

An experiment on hesitations or repairs will get too large to handle if you use more than about 6 to 8 participants, and it should be directed towards a fairly specific hypothesis. For classic references see:

Schegloff, E., Jefferson, G., and Sacks, H. (1977) The preference for self-correction in the organization of repair in conversation. *Language*, 53, 361-382.

Levelt, W.J.M. Monitoring and self-repair in speech. (1983) *Cognition*, 14, 41-104.

For other empirical studies of disfluencies (only the first is an experiment) see:

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, 32, 25-38.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, 62, 151-167.

Clark, H. H. & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speech. *Cognition*, 84, 73-111.

**Language Through Computers**

With the advent of email, instant messaging, chat rooms, texting, and Facebook, there are many issues you can examine about communication through computers and how this type of communication differs from other forms. This is new enough that there isn't a lot of research as background. The trick is to create two or more conditions for comparison—conditions that differ in interesting features—and create transcripts or logs of what people did.

(1) IM comes in several versions. From what you know about face-to-face conversation, you might have hypotheses about which of these versions is more efficient, or less prone to misunderstanding. You could also compare IM and F-to-F conversions on the same tasks, but here you want clear expectations about what to

look for and why.

(2) People use short cuts in IM and texting, including abbreviations, smileys, and other typographical inventions. You can compare the use of these conventions under various conditions—say, between friends vs. strangers, or between vs. within age cohorts. Again, be sure you have clear hypotheses about what you should find and why.

For one approach to comparing media, see Clark and Brennan (on my website). Crystal also has book called *TXTING*.

Clark, H. H., and Brennan, S. A. (1991). Grounding in communication. In L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.). *Perspectives on socially shared cognition*. Washington: APA Books.

Crystal, D. (2008). Texting. *ELT journal*, *62*(1), 77-83.

**Slips of the Tongue and Tongue Twisters**

It takes a long time to collect enough spontaneous slips of the tongue (e.g., spoonerisms, blends, etc.) to study them in detail, but you can induce them by several methods. One is the Spoonerisms of Laboratory Induced Predisposition (SLIP) technique devised by Motley, Camden, and Baars to look at various hypotheses about slips. For this, see Motley et al. Warning: It takes more data than you might think to get this technique to work. Another method is to devise tongue twisters with just the properties you want and have people repeat them, say, five times quickly (while you record them). See Gary Dell et al. and Stephanie Shattuck-Hufnagel. Here are two hypotheses that have been tested.

(1) Slips are more often words than nonwords. So using SLIP, you can show that it is easier to induce a slip that is a word (*dull fog* going to *full dog*) than one that is not a word (*dart board* going to *bart doard*).

(2) Slips occur more often between syllable initial consonants (e.g., *hip fan* going to *fip han*) than between syllable final consonants (*hip fan* going to *hin fap*). This was shown using carefully selected tongue twisters that people had to repeat over and over again quickly. Note: You do not need to use standard tongue twisters; you can make them up to suit your own purposes (as Dell always does).

You can test a variety of hypotheses with these techniques. (1) Are people more likely to anticipate than perseverate in slips of the tongue? (2) Are slips of the tongue more likely with longer than with shorter words? (3) Are people more likely to produce slips when they cannot hear (and therefore monitor) what they are saying, e.g., when they are wearing earphones and listening to loud rock music?

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, 104, 123-147. (For his more recent papers on slips, use Google Scholar.)

Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 21, 578-594.

Shattuck-Hufnagel, Stephanie. (1992). The role of word structure in segmental serial ordering. Special Issue: Lexical access in speech production. *Cognition*, 42(1-3), 213-259.

**Conversational Routines**

It is possible to gather spontaneous bits of conversation between clerks and customers just by sitting nearby and recording, by hand, what is said. To do this, you need specific hypotheses to test, specific patterns to look for.

(1) You might have a hypothesis about the brevity, or ellipsis, of the requests being made of the server depending on whether what they are asking for is routine or not. Or you might have hypotheses about the server's response depending on whether the customer's request was elliptical or not.

(2) A variation on this is to introduce an experimental variable into such situations. Take a simple example. If you ask someone for the time with "Can you tell me the time?" they will sometimes answer the literal question "yes" before going on to tell you the time. Are they as likely to answer the literal question with "Do you know what time it is?" and "Would you mind telling me what time it is?" and so on? You can test a variety of hypotheses with this sort of technique. For an example of its use, see:

Clark, H.H. (1979) Responding to indirect speech acts. *Cognitive Psychology*, 55, 797-811.

(3) Some expressions and conversational routines are more polite than others, and they get used by servers with some customers but not with others, and vice versa. You can study the choice of such expressions either experimentally (as in the previous suggestion) or merely by observation in various public circumstances. The opposite of politeness, of course, is rudeness, and you might compare situations in which servers or customers are ruder than those in which they are not. Start, however, with a clear hypothesis or rationale. *The* classic book on politeness is this:

Brown, P., & Levinson, S.C. (1987) *Politeness: Some universals in language usage*. Cambridge University Press.

(4) You can also look at turn-taking. In different situations, the transfer of turns may take different forms, and be accomplished by different means. What about, for example, peer and peer vs. boss and employee, or student and professor, or turns in chat sessions? Remember that it is illegal to record other people's conversations without their permission. For *the* classic, but chewy paper on turn taking, see:

> Sacks, H., Schegloff, E., and Jefferson, G. (1974) A simplest systematics for the organization of turn-taking in conversation. *Language*, 50, 696-735.

(5) Or you can look at patterns of collaboration in conversation. As we know, two people in conversation add "uh huh" in the background, overlap each others' speech, initiate side sequences for repairs, divide their utterances into installments, and complete each others' utterances as they collaborate with each other. Precisely how they do this and how often depends on the status of the two people, on their intimacy, on what they are trying to accomplish, etc. There are many hypotheses in this area that remain to be tested. See:

> Clark, H. H. & Schaefer, E. F. (1989) Contributing to discourse. *Cognitive Science*, 13, 259-294.

**Corpus research**

There are many issues you can study by counting words, phrases, and the like in corpora of spoken conversations, books, plays, you-name-it. Here are a few corpora, though there are many more now available:

- British National Corpus: available at: http://corpus.byu.edu/bnc (huge!)
- Other spoken and written corpora in American English: http://corpus.byu.edu/
- Books from 1800 on: https://books.google.com/ngrams (huge!)
- Switchboard corpus of telephone conversations: e.g., http://groups.inf.ed.ac.uk/switchboard/

You'll need time to figure out how to use these, but most have instructions. In some cases, you may need programming skills.

My students and I have used "A corpus of English conversation," edited by Svartvik and Quirk, because it includes intonation, speech errors, overlapping speech, etc., properties not found in most other corpora. We were able to investigate "the" pronounced as "thee," "uh" and "um," and repeated words, as in "I uh I think that we should go." (References to these papers are found earlier in this document.) The problem is that Svartvik-Quirk corpus is not public, as far as I know, though you can do some of the same studies with Switchboard.

So what can *you* do with these corpora? Almost anything that you can do by

counting words or phrases. But whatever your interest, you need to think statistically. You'll need to *compare* counts, or rates, of the items you look for, and you'll need to control for obvious confounding factors. Some versions of Switchboard have the durations of words, turns, and pauses, but investigating durations would take some programming skill—and some worthy hypotheses about what to compare.

### Julie Sedivy's *Language in Mind*

The textbook by Julie Sedivy has a website full of demonstration experiments of all kinds, classified by chapter. You can pick a topic you are interested in, find the right chapter, and look through the demos she has collected for that chapter. Here is the website:

http://sites.sinauer.com/languageinmind/index.html

Warning. It would be trivial to expand on and run the experiments she has suggested, but I am challenging you to do something more. Design and carry out your own experiment, even if you make use of a method she has suggested.

Good luck!