# What Should We Do With Our Speech Error Corpora? Notes from the Panel Discussion

*Carson T. Schütze*
*University of California, Los Angeles*

*Victor S. Ferreira*
*University of California, San Diego*

## 1. Introduction

We summarize in this paper some of the issues raised, opinions voiced, and ideas put forth at the eponymous panel discussion, for which we served as moderators. The panelists were Gary Dell, Jeri Jaeger, Stefanie Shattuck-Hufnagel and Joe Stemberger, whom we would like to thank for their contributions, and there was considerable audience participation as well. For practical reasons we do not attempt to attribute comments, nor to preserve the sequence of topics discussed. For further discussion of some of the issues touched upon herein, see also Shattuck-Hufnagel 2007.

### 1.1 Historical background

Much of the early work relating speech errors (and thus language production) to linguistic structure was based largely or entirely on corpora of naturally-occurring slips that were written down by researchers in the course of their daily activities. Many large corpora consisting of thousands of slips each were amassed in the 1970s and 1980s. It did not take long for psycholinguists to be aware of the potential problem of observer bias in these corpora: What if the errors that people write down are a non-random sample of all the actually occurring errors? What if certain error types are systematically over- or under-represented? One way of addressing this problem was to develop experimental techniques for eliciting particular kinds of slips; these of course raise their own issues of ecological validity. Nonetheless, one could at least ask to what extent the patterns in elicited vs. corpus-based slip data were consistent with each other, and a few such comparisons were reported on, though they unfortunately did not agree on the answer to this question. (From today's perspective it seems likely that the answer might differ depending on the specific empirical issue of interest: for instance, the relative frequencies of different slip types might be expected to show more task dependency than the degree of similarity between units involved in a particular slip type.)

In the 1990s and 2000s, language production research adopted a widening range of methods. However, many researchers still believe that speech error data provide unique and crucial insights into the workings of the production system,

*MIT Working Papers in Linguistics 53, 383–393*
*The State of the Art in Speech Error Research: Proceedings of the LSA Institute Workshop*
Carson T. Schütze and Victor S. Ferreira (Eds.)

at least in principle. In addition, even if we could be certain that elicited errors perfectly mirror spontaneous errors, there are many types of slips for which no elicitation paradigm has been developed, and prospects for these may be slim (e.g. shifts, blends, multi-word errors, etc.). Furthermore, slips produced by populations such as children or people with language disorders arguably are not amenable to study by (m)any of the alternative methods applicable to the typical adult population.

One solution that ought to allow a return to the use of spontaneous errors is to cull them from extended passages of recorded speech, carefully transcribed by multiple researchers during repeated listenings. There remain some problems in principle with this method, such as the fact that some slips might go unnoticed even on repeated listenings although they are perceptible, and other slips that might have been detectable in face-to-face conversation may not be perceptible or identifiable from recordings due to reduced access to relevant context.[1] (For example, unless you restrict yourself to recordings of speech situations you yourself participated in, your chances of knowing what the speaker intended to say will on the whole be reduced.) Also, the opportunity of asking a speaker what he or she meant to say is obviously lost. (For further discussion of these and related issues, see Cutler 1982a; MacKay and Kempler 1984.) What most people seem to perceive as the bigger problem with transcription from recordings, however, is that it is extraordinarily time-consuming and hence expensive, owing in part to the sparseness of slips, particularly in the sort of speech for which it is easiest to find high-quality recordings (TV and radio broadcasts, for example).

Where does this leave us? Specifically, are corpora of spontaneously-observed errors still useful in language production research? The above considerations lead us to think that they are, though this point is evidently at least a little controversial.

### 1.2 Motivation

In this milieu, the idea for this panel arose from a very practical concern: What should be done with the UCLA Speech Error Corpus (UCLASEC), which contains about 3,000 errors (collected mainly by Vicki Fromkin), once the software that had been set up to manipulate it as a database was no longer supported on current-generation computer hardware? Since people continue to request copies of the corpus, converting it to a modern, platform-independent form seemed appropriate, so using the World Wide Web was the obvious approach. The question then arose as to whether the goal should simply be to duplicate the contents and functionality of the previous database system, or whether this might be an opportunity to make improvements of some sort. Going to such effort seemed more appealing if it could yield benefits beyond just this one corpus. Given the existence of at least a handful of slip corpora of similar size at other institutions, could a system be designed that would, after a manageable amount of conversion effort, be able to store and search these other corpora in a uniform way, ef-

[1] Recordings also fare no better in general than spontaneous corpora or traditional elicitation data with regard to "errors" that are not acoustically detectable (by the human ear or at all); these seem to call for special (phonetic) experiments in any case.

fectively creating a super-corpus that might open up new research possibilities? (The CHILDES Database (MacWhinney 2000) was an obvious inspiration.)

This is the context in which many of the issues discussed below originally arose. It was not known whether the research community would see value and viability in centralizing and standardizing the various corpora of speech errors that are in (or could be placed in) the public domain, and making them searchable via a uniform web-based interface. Even so, there would be at least two fairly clear benefits. One would be to open up the use of speech error data to a much broader research (and teaching) community, i.e., anyone with access to the Web. Another would be to mitigate methodological concerns about bias and lack of representativeness in existing individual corpora, to the extent that different collectors have different biases that could average out. But would it be worth the effort, and if so, how should it be done?

## 2. What to do with existing speech error corpora

In preparing for the panel we laid out for discussion a number of issues that would need to be addressed before embarking on the project envisioned above. Not all of these were actually discussed in the 90-minute session, but we felt some of them were still worth thinking about, so they are preserved here. As a result, in what follows there will be numerous questions to which no answers are suggested. We have also interjected some of our own responses that we did not voice during the panel. Finally, we acknowledge that helpful and provocative things were said by many people in the room that, for various reasons, are not mentioned here.

### 2.1 How ambitious should we be?

The central question driving the discussion was: What is the best way to take advantage of all the effort that has already been put into amassing corpora of spontaneous speech errors? One possible answer is that nothing needs to be done: people who want to use the corpora are finding ways to do so (as witnessed by the corpus-derived results in several of the papers in this volume), and there is no pressing demand to make them more accessible or improve their functionality. Although it may be true that there are no hoards knocking at the doors of corpus-holders clamoring for access to the data, this may reflect a chicken-and-egg situation. The common assumption may be that the corpora are too difficult to use, even if their owners are willing to share them—perhaps they are still stored on index cards, or they are completely uncoded, etc., so that there would be no point in looking at them.

A different answer to the question could be this: Today's technology would allow many more people to take advantage of these resources than are currently doing so, if someone is willing to invest even just a little effort. For example, might it be useful to simply take all corpora that are in computer files, make those files downloadable over the web, and create a central website ("Slip Central"?[2]) with labeled links to all the download sites. It would be up to the user to deal with whatever format a given corpus happens to be in, something

[2] For those who prefer acronyms we offer ROSEC—"Registry of Speech Error Corpora."

that its provider could choose to provide help with or not (e.g. by way of a list of data fields, coding conventions, etc.).

From this rudimentary concept, one can envision various degrees of elaboration. For example, add to the above a summary chart on Slip Central listing such information as which language(s) the errors are in, how many there are, whether they were restricted to a particular type (e.g. only phonological ones), who the speakers are (native/nonnative, adults/children/clinical populations), how the observations were gathered (all by the same observer vs. from students as part of a class assignment), and so on. This remains within the bounds of a model on which the corpus holders are fully responsible for what is actually available to users, and Slip Central merely serves the function of helping users find corpora that may be of interest to them. In the discussion, the majority opinion seemed to be that this was worth doing.[3]

A much more ambitious approach would be for Slip Central to impose some requirements on the corpora that it lists, such that users who have figured out how to use one corpus for their purposes can transfer some of that knowledge to another corpus. Minimally this might consist of agreeing on a file type (ASCII text, SQL, XML, or something else), which would obviously have implications for the way errors would be laid out in these files (e.g., a row of a spreadsheet represents one error, the columns represent different pieces of information about it; or delimited text with unique identifiers for the start and end of each error, etc.). The next step would be to agree on the format in which certain descriptive information would appear for each error (e.g., erroneous material is boldfaced, Unicode IPA is used for any transcriptions, …). Since software already exists for converting between various ways of representing database/spreadsheet contents, getting this far might still not require much manual labor—some scripts might need to be written to automate the conversion process.

Going further, one might try to standardize some of the contents, not just the format, of the corpora. A model that was suggested for this approach is the CHILDES Database, in which all contributed corpora must follow a standard, computer-enforced scheme for coding the large variety of types of information that can be included in them. This has allowed for the development of a general set of software routines for manipulating and analyzing any corpus in this format. The sentiment at the workshop seemed to be that this was probably not feasible in the domain of speech errors, for two basic reasons: it would be impossible to agree on what the standard format should be, and it would require too much work to convert existing corpora from their current form to some new scheme. (On the latter point, hopes for creating automated conversion algorithms, even if they were unique to a particular existing corpus, did not seem bright.)

Nevertheless, let us play out this fantasy one episode further. On analogy to what LDC provides for searching the Treebank corpus (Marcus, Santorini and Marcinkiewicz 1993), if corpora were standardized as above we could in principle gather (shadow copies of) them at the Slip Central site and build a search in-

[3] We are putting aside the issue of what to do with corpora that have not been put on computer in any form—presumably the motivation (and the funding) for this huge task would have to stem from one's own research goals.

terface that could be used from a web browser (using pull-down menus etc.) to search all of them at once (or a specified subset, e.g. all those in a given language), effectively creating a mega-corpus at least an order of magnitude larger than any that now exist.

### 2.2 Potential obstacles

Even the least ambitious program discussed in this section presupposes the willingness of those who currently control slip corpora to make them publicly accessible (as UCLASEC has been). Is this likely? Two obvious kinds of hindrance may arise. First, researchers may feel that their corpus is "not ready for prime time," that is, they don't want to make it public without fixing/cleaning it up in certain respects, and who knows when they will get around to doing so. Second, they may be hesitant to share a resource that cost them considerable time, money and effort to create and which provides them with unique research/publication opportunities. Participants in the discussion felt that neither of these should be allowed to prevent most corpora from being put in the public domain. Experience suggests that access to an imperfect corpus is still much more useful for research than no access at all (particularly if the nature of the flaws is known and can be documented). As for proprietary inclinations, the perception is that there is considerable moral pressure on researchers who use resources such as CHILDES to contribute their own data in return, possibly after a relatively short period during which the corpus collector gets first crack at it.

There may also be legal and ethical concerns. It was noted that there is a problem with corpora that have been published in books, where the publisher may hold the copyright and not be willing to allow an electronic version to be freely distributed. There was some discussion about whether making *any* spontaneous speech errors publicly accessible would violate privacy laws or confidentiality requirements or would require consent from the original speaker, given that what one says by its nature often includes self-identifying information. Many people believed that spontaneous slips are drawn from publicly observable behavior and therefore not subject to these requirements, while others wondered whether it was reasonable to claim that any time someone can hear another person talk that person is publicly observable. It was also pointed out that speech collected from clinical populations might be subject to unique restrictions.

### 2.3 What could be the common format?

Considerable discussion was devoted to the issue of how different researcher's corpora could be standardized so that, for example, one could do an automated search for all the blend errors in all the English corpora. The conclusion was that this is impossible, for two reasons: people do not agree on what information should be in an error record, and they do not agree on how to code errors for purposes of analysis.

Above we spoke as if a core set of descriptive parameters would clearly be agreeable to everyone, but even that seems not to be true. For example, isn't it simple enough to say that every record should contain one field for what was actually uttered and one field for what was intended to be uttered? Points of contention arise already here: What if there are two possibilities for what was in-

tended—to what extent will the values of other fields depend on which one is assumed, and how will each set of values be linked to just one of the targets? Does this require two separate records? If so, how do we represent the fact that these were not two separately-produced errors? We need to do so in such a way that when we write routines to count errors they do not count this one twice. (Similar issues arise when a speaker produces more than one incorrect version of the target, but then it is not obvious whether this utterance should count as containing one error or two.) What if part of the target could not be determined at all—should such errors be specially flagged? Should they be omitted from the database? Should we record a distinction between errors in which the target was determined from the speaker's own self-correction versus those where it was merely inferred by the listener (the latter assumed to be less reliable)? All this, and we are still dealing with the two most "straightforward" descriptive data fields!

With some pieces of descriptive information it would be fairly easy to say that if anyone wants to include it in the common format there will be a field for it,[4] but researchers who did not record that information with their slips are free to leave it blank. This could include things such as who the speaker was, who observed the error, the date, the linguistic nature of the situation, and so on, though a given researcher might actually consider one of these pieces of information vital and wish to ignore errors for which it is unavailable. An alternative is simply to have a "Comments" field of effectively unlimited size where the researcher can record any of this information that they wish to. This makes life easier for everyone except the person who someday wants to partition the corpora based on one of these criteria; e.g., it could be interesting to compare the distribution of error types among all errors observed by, say Vicki Fromkin versus those observed by Merrill Garrett (both are well-attested in the current UCLASEC database), to check for idiosyncratic biases in which errors they took note of.

All of these problems pale in comparison to those involving precisely the aspect of the corpora that makes them prima facie most appealing to the research community: the classification of the errors. We may well be able to reach agreement on what the set of possible error types is; the problem arises in specifying criteria for being of a particular type. Panel members noted that they have written multi-page guidelines to try to keep this consistent within their own corpora, but certainly the community will not agree on one set of guidelines for everyone—how to classify an error depends to some extent on your (implicit or explicit) theory of precisely how errors arise, as well as how you intend to use the corpus.

A lot of discussion on the panel centered on the issue of how to treat errors that have more than one possible analysis—ambiguous errors. Here a diversity of approaches was apparent. In some cases, new error categories have been created just for situations that would otherwise have alternate analyses, e.g. the situation where a target slips in a context that is both preceded and followed by

[4] The ensuing plethora of data fields can become unwieldy and complicate search interfaces, however. This is evident in the UCLASEC system, which has about 70 fields—they cannot all usefully be displayed on a computer screen at one time, a source of ongoing frustration for both users and programmers.

something relevantly similar to the erroneous output (i.e., an anticipation/ perseveration). This might embody a theoretical claim that the combination of the two contextual influences is more than just the sum of each individual one. In other situations where two analyses are in principle possible, researchers include in their coding guidelines a rule dictating which one analysis should be used. For example, in sound errors a great many cases where a single phoneme is affected could also be described as one phonological feature changing. Many researchers do not code both of these analyses, only the phoneme-based one. In this case the reason is that if one compares the frequency of errors that are unambiguously in one category versus the other, phoneme errors are much more frequent than feature errors.[5]

These two approaches will not cover all ambiguous errors, however. Most speech error researchers agree that a certain portion of a slip corpus (the overwhelming majority, according to some) will have to indicate multiple possible analyses for a given slip. Above we noted, in the context of alternative target utterances, some of the technical problems that arise: how do you search and count such errors properly? Specifically, what if a given slip is, say, a phoneme error on one analysis and a morpheme error on another analysis—when we ask our computer how many phoneme errors there are in the corpus, what answer do we want it to give us? The web-based version of UCLASEC currently under development sidesteps the question by reporting an upper and a lower bound for all its counts. That is, it will tell us "There are certainly at least *n* of what you were searching for, and there could be as many as *n+m*, depending on how the ambiguous errors fall."[6] That may be the most neutral answer we can expect, but such ranges do not lend themselves well to subsequent calculations comparing numbers of different error types.

Beyond statistical challenges lies a harder question: how does one decide whether a particular error should be treated as ambiguous, and which and how many of its possible classifications should actually be listed in a coded corpus? Pre-specified decision criteria are not so useful here, because the researcher's intuition seems to be that the likelihood of a particular slip-causing scenario is not an absolute, but rather is relative to what the alternative possibilities are and how likely they are. That is, an analysis in a context with a much more plausible alternative may not be coded for, but the very same analysis in a context where the only alternative is of similar likelihood will be. We may go pretty far in order to avoid having to say that we have no idea how an error arose. The whole discussion presupposes that there is some way to assign (at least relative) probability to slip scenarios, but beyond the most basic cases it is not obvious that people are even aware of how their own intuitions arise. Moreover, researchers may differ on which slip scenarios even occur to them as possibilities, and it does not

[5] Statistically speaking it would be more logical to assign the ambiguous errors to the two categories in the same proportions that are observed among the unambiguous errors. Practically speaking, however, this approach does not tell you what to do with each particular error as you come to code it, and it requires coding a suitably large sample of unambiguous errors first.

[6] The intuitively correct lower bound is not simply the number of unambiguous errors of the searched-for type. Rather, it should also include ambiguous errors all of whose analyses satisfy the search criteria.

seem likely that we can remedy this by the standard route of creating a checklist of all possible analyses to be considered—the combinatorics are too vast.

The conclusions from this discussion are the following: Slip Central should facilitate locating and understanding the contents of slip corpora but should not try to regulate what the contents should be. In the end, serious researchers will end up reading through all the errors themselves anyway, devising coding schemes that suit their purpose at hand.

## 3. Moving beyond existing corpora

### 3.1 New spontaneous data?

Is there any point in continuing to write down the slips we hear in everyday situations? Given what we know now, could we collect this kind of data in a way that would address some of problematic aspects? For example, does requiring a corroborating observer, as Vicki Fromkin did (Fromkin 1973, p. 112), actually improve the overall reliability of the corpus or does it reduce the sample size in a potentially biased way? Could modern digital microrecording technology provide a convenient way of implementing an early attempt to try to ensure that we keep a record of every slip we notice and to circumvent the inaccuracy of observer memory? The idea would be to keep a digital recorder running continuously during one's daily activity; the recorder would retain the past minute's worth (say) of audio and erase earlier material, but when the user presses a button it would copy the current one-minute segment to a longer-term memory for later analysis.[7]

### 3.2 Slips transcribed from recordings

Above we mentioned the pros and cons of collecting slips "exhaustively" by repeated listening to recorded speech. Given that some researchers are undertaking this resource-intensive task, would it not make sense to make this kind of data shareable too? Might this in fact be a higher priority than making corpora of spontaneously-observed slips available? Would it be important for the entire contents of the recording to be transcribed,[8] or just the errors and their immediate context (and how much is "immediate"?)? Would it be important for the (audio or video) recordings themselves to be available in an archive along with the transcribed errors? One might want to track additional information about the recording as a whole in this kind of a slip collection, e.g. duration of speech and total number of words/syllables/etc. uttered by each of the speakers, speech rate within a given window surrounding the slip, and so on.

---

[7] The original version of this procedure, probably due to Anne Cutler, employed tapes that would be swapped out of a continuously-running tape recorder as soon as an error was noticed. That is, each "used" tape would contain one slip, just before where the recording stopped.

[8] This could make it possible to calculate base rates for (potential) error contexts that are much better tuned to the speaker, in contrast to current practice, which tends to have to rely on unrelated language samples. (E.g., one could count how many occurrences of each lexical category were produced and then calculate the slip rate within each one.) The trade-off as usual is a decrease in sample size.

It was noted that there are already some corpora available that consist of recorded speech and a transcription thereof, and so perhaps we could take advantage of those and save ourselves a lot of effort. The question that arises is to what extent any slips that were produced on the recording may have been (consciously or subconsciously) corrected by the transcriber, since finding slips was not their express purpose at the time. It would be terrific if we could persuade the people who are creating these transcripts to flag slips in some uniquely-searchable way.

There was a little discussion of what kind of speech situations we should try to record, if finding slips is our primary goal. Suggestions included academic classroom lectures, which are typically unscripted but often contain complex sentence structures, low-frequency words, etc., and live TV/radio news reporting ("breaking news"), where reporters may have to describe unexpected events in real time as they unfold, but must also prevent "dead air" by continuing to talk even when there are no new events to describe.

### 3.3 Other sources of slips

There are by now a handful of different ways to elicit slips of the tongue in the laboratory, ranging from versions of the SLIP[9] paradigm (Motley and Baars 1976) and tongue-twisters (MacKay 1971; Shattuck-Hufnagel 1983) to much more open-ended techniques (e.g. Dollaghan et al. 1990). If the effort is made to develop a centralized system for accessing spontaneous slip data via the web, we ought to ask whether experimentally-elicited slips would be usefully stored in the same kind of system. Participants seemed to think this was worth exploring, given that even in such experiments the density of slips is not very high. At the same time, the limitations of these paradigms are felt by some people to be quite serious.[10] Indeed, we were reminded that all of our empirical data are flawed in one way or another, so this by itself cannot justify throwing out one particular kind. The question should always be, "Does this method provide useful information that cannot be obtained by other available methods?" The panel discussion made explicit what was implicit in the rest of the workshop: many people still answer "Yes" with regard to speech error data, even if they admit to having

[9] An acronym for Spoonerisms of Laboratory-Induced Predisposition.

[10] The field is clearly looking for new methods of slip elicitation that maximize the number of slips that occur while minimizing the artificiality of the task. At the panel, Schütze noted that there is a drug prescribed for migraine and depression that has the well-known side effect of disrupting one's everyday speech, by reducing fluency and/or increasing errors. It seems reasonable to assume that whatever this drug is doing is probably not changing the nature of people's slips very much: for example, it is unlikely to be increasing the relative proportion of stranding errors.

Schütze had off-handedly remarked to a doctor that it was too bad that medical researchers studying migraineurs and people with depression were unlikely to be interested in supporting the relatively frivolous endeavors of psycholinguists. However, the doctor noted that this drug has proven to be so well-tolerated (except when it hinders your communication) that it is now expressly prescribed for what was originally just one of its side effects, namely weight loss. Since there is an immense population of candidates for weight loss drugs, many of whom seem eager to try an impressive range of purported treatments, it might not be hard to find volunteers for a "Slip 'n' Slim" study.

themselves abandoned it for the particular research issue occupying them at the moment.

### 3.4 Other kinds of production data

Once the idea of a central web site for speech error data is firmly established, we should consider what else could usefully be put on such a site. An obvious suggestion, again inspired by CHILDES, is a bibliography of research involving slips of the tongue. In fact, a modern electronic version of such a database, founded on Cutler's (1982b) printed list, has been assembled at UCLA and currently contains almost 900 references. Obviously it will require the participation of the research community to keep it (reasonably) complete and up-to-date.

One might also consider whether other sorts of data collections might be useful to the same research audience as slips. The prime candidate would be tip-of-the-tongue (TOT) data (for which a bibliographic database has also been gathered at UCLA, containing 160 references). Parallel to the situation with slips, one could assemble both spontaneous and elicited TOTs. Contrariwise, since the task for inducing the TOT state is quite standard (Brown and McNeill 1966), it is not absurd to consider doing quantitative analysis on TOTs collected across multiple experiments in different labs. It would also seem much easier to establish consensus on what pieces of information should be stored in a TOT record, since again the research literature is fairly uniform in its dependent measures. Another major data source for language production are hesitation phenomena, but these seem closer to slips than to TOTs in terms of the complexity of coding issues they raise.

As we add new methodological tools to our research arsenal, we should continually ask ourselves whether the resulting data could usefully be shared with others in the spirit of scientific cooperation.

## References

Brown, Roger and David McNeill (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior* 5, 325–337.

Cutler, Anne (1982a). The reliability of speech error data. In *Slips of the tongue and language production*, ed. Anne Cutler, 7–28. Berlin: Mouton.

Cutler, Anne (1982b). *Speech errors: A classified bibliography.* Bloomington, IN: Indiana University Linguistics Club.

Dollaghan, Christine A., Thomas F. Campbell and Russel Tomlin (1990). Video narration as a language sampling context. *Journal of Speech and Hearing Disorders* 55, 582–590.

Fromkin, Victoria A. (1973). Slips of the tongue. *Scientific American* 229(6), 110–117.

MacWhinney, Brian (2000). *The CHILDES project: Tools for analyzing talk.* 3rd edition. Mahwah, NJ: Lawrence Erlbaum.

MacKay, Donald G. (1971). Stress pre-entry in motor systems. *American Journal of Psychology* 84, 21–43.

MacKay, Donald G. and Daniel Kempler (1984). Review of *Slips of the tongue and language production*, ed. Anne Cutler, Berlin: Mouton, 1982. *Language* 60, 436–441.

Marcus, Mitchell P., Beatrice Santorini and Mary Ann Marcinkiewicz (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19, 313–330.

Motley, Michael T. and Bernard J. Baars (1976). Laboratory induction of verbal slips: A new method for psycholinguistic research. *Communication Quarterly* 24, 28–34.
Shattuck-Hufnagel, Stefanie (1983). Sublexical units and suprasegmental structure in speech production planning. In *The production of speech*, ed. Peter F. MacNeilage, 109–136. New York: Springer.
Shattuck-Hufnagel, Stefanie (2007). What traditional speech error corpora can tell us (and what they can't): How not to throw the baby out with the bathwater. This volume.

(Schütze)
Linguistics Department
UCLA
3125 Campbell Hall
Box 951543
Los Angeles, CA 90095-1543
USA

*cschutze@ucla.edu*
*http://www.linguistics.ucla.edu/people/cschutze*

(Ferreira)
Psychology Department
UCSD
9500 Gilman Drive
La Jolla, CA 92093-0109
USA

*ferreira@psy.ucsd.edu*
*http://lpl.ucsd.edu/*