Toward a Taxonomy of Communication Errors

Tim Paek

Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
timpaek@microsoft.edu

Abstract

Researchers in a variety of disciples such as conversation analysis, second language acquisition, speech pathology, and computational linguistics have classified various types of communication errors that occur between humans, and between humans and computers. In this paper, we attempt to bring together insights from diverse disciplines into a common theoretical framework from which to compare and contrast various types of errors. We describe a classification scheme based on how people collaboratively resolve uncertainties by establishing the mutual belief that their utterances have been understood well enough for current purposes—a process referred to as *grounding*. The classification scheme not only highlights how different types of failures are related to each other through the common bond of uncertainty, but also provides guidelines for the design of error handling in automated systems.

1. Introduction

When people engage in a conversation, they typically do so with the intent of making themselves understood. As such, they need to make sure that the other participants are attending to them, actively listening, and understanding what they are saying. Since unresolved uncertainties can lead to communication errors, people cooperate with each other to establish and maintain the mutual belief that their utterances have been understood well enough for current purposes ([6], [7], [8]). An example of this kind of collaboration is the use of backchannels. When participants in a conversation sufficiently understand an utterance, they will typically give feedback through head nods or acknowledgements, such as "uh huh". On the other hand, if they do not understand, they will attempt to clear up their uncertainties until they have coordinated not only the content of their utterance but also their beliefs about what they mutually understand ([6]).

Researchers have described this kind of elaborate coordination as a type of "joint activity" ([6], [9], [13], [27]). Just as a dance is more than the sum of individual motions, a conversation is more than a structured sequence of utterances. The process by which participants coordinate the presentation and acceptance of their utterances to establish, maintain, and confirm mutual understanding has been called *grounding* ([6], [7], [8]).

Although researchers in a variety of disciples such as conversation analysis, second language acquisition, speech pathology, and computational linguistics have classified various types of communication errors that occur between humans, and between humans and computers, little effort has been made to compare and contrast these types across disciplines. Taking *grounding* as a theoretical framework, we propound a classification scheme for examining how different types of communication errors are related to each other through the common bond of uncertainty. In the first part of this paper, we describe the classification scheme and delineate an interdisciplinary taxonomy of communication error types. We also discuss how the scheme can be utilized for error handling in automated systems. In the second part, we explore particular error types identified in fields outside of computer science that offer unique insights into designing repair strategies for error recovery.

2. Classification scheme

When people engage in a conversation, they make sure that the other participants are at the same time attending to, hearing, and understanding what they are saying. Following Clark ([6]), speakers and listeners ground mutual understanding at four levels of coordination, as shown in Table 1.

Table 1: Four levels of coordination for grounding mutual understanding.

Level	**Speaker *S***	**Listener *L***
Conversation	*S* is proposing activity α	*L* is considering proposal α
Intention	*S* is signaling that *p*	*L* is recognizing that *p*
Signal	*S* is presenting signal σ	*L* is identifying signal σ
Channel	*S* is executing behavior β	*L* is attending to behavior β

At the most basic level, the *channel* level, a speaker *S* attempts to open a channel of communication by executing behavior β, such as an utterance or action, for listener *L*. However, *S* cannot get *L* to perceive β without coordination: *L* must be attending to and perceiving β *precisely* as *S* is executing it.

At the next higher level, the *signal* level, *S* presents σ as a signal to *L*. Not all behaviors are meant to be signals. For example, scratching an itch during a conversation is usually irrelevant to interaction at hand. Hence, *S* and *L* must coordinate what *S* presents with what *L* identifies.

The *intention level* is where people interpret the semantic content of signals. At this level, *S* signals some proposition *p* for *L*. What *L* recognizes to be the intention of *S* in signaling σ is how *L* will arrive at *p*. Note that the signal σ is different from the *intention* of *S* in using σ (e.g., in indirect speech acts). *S* cannot convey *p* through σ without *L* *recognizing* that *S* intends to use σ. This again takes coordination.

Table 2: An interdisciplinary taxonomy of communication error types.

Channel	Signal	Intention	Conversation	Communication Error Types
1	1	1	1	Channel failure, overhearing (Paek & Horvitz, 2000)
0	1	1	1	"Non-understanding" (Hirst et al., 1994); "Problematic reference" (Schegloff, 1987); "Communication breakdown" by cochlear-implant users (Tye-Murray & Witt, 1996); Aphasic "jargon" (Ahlsen, 1993)
0	0	1	1	"Misunderstanding" (Hirst et al., 1994) by abduction (McRoy & Hirst, 1995) by coherence (Ardissono et al., 1998); "Negotiated misunderstanding" (Blum-Kulka & Weizman, 1988); "Conversational breakdown" in cross-cultural interaction (Ulichny, 1997); "Input failure" (Ringle & Bruce, 1981); Underspecification (Deemter & Peters, 1996)
0	0	0	1	"Misconception" (McCoy, 1989); Breakdown by "unshared conventions" (Gumperz, 1995); Co-membership differences (Schegloff, 1987); "Problematic sequential implicativeness" (Schegloff, 1987); "Pragmatic infelicities" (Marcu & Hirst, 1996); "Model failure" (Ringle & Bruce, 1981)
0	0	0	0	"Contribution" (Clark, 1996)

Finally, at the *conversation level*, *S* proposes some joint activity *α* which *L* considers and takes up. A proposal solicits an expected response defined by *α*. For example, in an indirect speech act such as "This room is cold," which is intended to mean "Please close the window," *S* is proposing an activity for *S* and *L* to carry out jointly—namely, that *S* gets *L* to close the window. *S* cannot get *L* to engage in the activity without the coordinated participation of *L* in closing the window.

2.1. Communication errors and uncertainty

The four levels provide a theoretical framework from which to examine the cause and effect of communication errors. The four levels form a set of co-temporal actions that begin and end together such that it is only possible to complete actions from the bottom level up. In other words, when speaker *S* executes *β*, it is done *in order* to present *σ*, which is done *in order* to signal that *p*, which is done *in order* to propose *α*. Clark ([6]) refers to this property as "upward causality." Since each level is built on top of its lower level, when a communication error occurs at a level, none of its higher levels can be complete. Hence, if *S* produces an utterance for *L* but *L* is not paying attention, then *L* will also fail to identify the signal and understand the intended meaning of the utterance.

Since speakers can be uncertain about whether mutual understanding has been achieved at any of the four levels, listeners try to give evidence that each level has been grounded through, for instance, feedback. In addition to upward causality, another property of the four levels is that evidence indicating that one level is complete constitutes evidence that all levels below it are also complete. Clark ([6]) calls this "downward evidence." For example, if *S* asks a question and *L* responds with an appropriate response, that is evidence not only of the fact that *L* understood the *α* that was being proposed, but also that *L* recognized *p*, identified *σ*, and attended to *β*.

Insufficient evidence at any level breeds uncertainty. When that uncertainty is enough to cause people in a conversation to engage in repair, they use upward causality and downward evidence to re-establish mutual understanding of their utterances. For example, when listeners are uncertain about whether they understood the intended meaning of an utterance at the intention level, they may question whether they heard the speaker correctly at the signal level. Or, when speakers are uncertain about whether a listener paid attention to their utterance at the channel level, due perhaps to lack of sufficient feedback, they may rightly wonder about whether the signal, intention, and conversation levels were jointly grounded as well.

2.2. Taxonomy

By specifying where uncertainty prevails in any of the four levels, it is possible to classify communication errors that have been identified in diverse disciplines such as conversation analysis, second language acquisition, speech pathology, and computational linguistics. An interdisciplinary taxonomy of communication errors is shown in Table 2. The taxonomy is not meant to be complete or exhaustive of the research literature; rather, it serves as a preliminary attempt to correlate different types of communication errors. The taxonomy was created by simply indicating in binary fashion whether any of the four levels needs to be grounded, or alternatively, whether unresolved uncertainty exists at any of the levels.

Although combinatorics dictates that with four levels, represented here as an ordered 4-tuple {channel, signal, intention, conversation}, a binary classification should yield sixteen (2^4) possible configurations, only five configurations are possible by upward causality and downward evidence. For example, if uncertainty exists at the intention level, then there must be uncertainty at the conversation level since the conversation level builds upon the intention level (i.e., {0,0,1,1}). Likewise, it may be the case that uncertainty at the intention level is caused by uncertainty at lower levels (i.e., {0,1,1,1} or {1,1,1,1}). There cannot, however, be uncertainty at the channel level with mutual understanding at higher levels. For example, in the case of {1,0,0,0}, it does not make sense for a listener to be inattentive to the behavior produced by a speaker, yet somehow understand what the speaker intended, with both listener and speaker believing that. Notice that when there is no uncertainty about joint understanding at any level (i.e., {0,0,0,0}), the utterance has been fully grounded and a contribution has been made to the conversation.

The taxonomy shows how various types of communication errors are related to each other. For example, what computational linguists have defined as a "non-understanding:" the failure to find a complete, unique interpretation of an utterance, is also commonly experienced in conversation with cochlear-implant users ([29]) and aphasic patients ([1]). This opens up the design possibility of using compensatory strategies that the hearing impaired employ in responding to errors, as we will discuss in the next section.

Another interesting relation is between a "conversation breakdown" in cross-cultural interactions ([30]) and a "misunderstanding" as examined by computational linguists working within the planning paradigm ([10], [15], [20]). To detect when a misunderstanding has occurred, planning-based systems typically formalize and exploit discourse expectations as constraints on user behavior. When a non-expected action is observed at the conversation level, the system reasons about whether this is a communication error at the intention level. Similar to these systems, when non-native speakers conversing with a native speaker unwittingly produce an expected response at the conversation level, they realize that downward evidence commits them to mutual understanding at the intention level, when in fact they do not comprehend. This is discussed in more detail in the next section.

By noting where people of various language impairments experience the need for grounding, the taxonomy is similar to examples given by Perlis et al. ([22]) of error type personas. The personas they present are as follows:

- An "apprentice" who is often lost in unfamiliar terminology and inferences and asks for clarification,
- A "child" who is often lost in "adult conversation" but uses social-cognitive skills to learn.
- A "foreigner" who has "no grasp of a language" but still engages in conversation.
- A person suffering from "SLI" (specific language impairment) who has "considerable difficulty" with ordinary spoken communication.

The purpose of the personas is to provide examples of conversational participants who experience language difficulty of some kind or another, but overcome them by relying on meta-dialog reasoning.

While Table 2 relates error types in a coarse fashion, finer-grained distinctions between types can be distinguished by applying the classification scheme to speakers and listeners separately as two groups. In fact, Table 2 really displays the union of the two groups. For example, in Table 2, a "misunderstanding," according to Hirst et al. ([15]), is when listeners believe they have a "complete and correct interpretation" of an utterance, but in fact do not have the one the speaker intended. In other words, listeners believe {0,0,0,0}, whereas speakers believe {0,0,1,1}. Since people must reach the mutual belief that their utterances have been understood well enough for current purposes, grounding must occur at {0,0,1,1}, or the union.

At this point, it is important to note that what is not represented in Table 2 are "communication errors" that remain under the surface. For example, Blum-Kulka & Weizman ([4]) classify "non-negotiated misunderstandings" in which speakers choose not to show that their intentions were misinterpreted. Instead, they choose to accept whatever interpretation the listener gave to their utterance. In other words, similar to a "misunderstanding" for Hirst et al. ([15]), speakers believe {0,0,1,1}, but decide to act as if they believe {0,0,0,0}. In such a case, it is unclear whether a communication error did in fact occur since the speaker has chosen not to treat the misinterpretation as an error.

2.3. Design implications

The classification scheme can be utilized for error handling in automated systems in two respects: first, by outlining an evaluative framework from which to compare the way automated systems encounter communication errors to the way their human counterparts do, and second, by creating a mapping of error types to repair strategies. With respect to the former, as explained previously, upward causality and downward evidence prevent certain types of the communication errors from occurring in human conversation. This is not to say, however, that they could not happen in an automated system ([21]). Indeed, since most spoken dialog systems are modular in integrating disparate black-box components such as the speech recognizer and template matcher, uncertainty can prevail in any of the levels without its effects being propagated to other levels. This presents the system designer though with a valuable comparative measure: Make sure that evidence is shared downward, and understanding is built upward so that whatever communication errors arise, they at least develop in such a way that people are accustomed to collaboratively dealing with them ([21]).

The second design implication is that the classification scheme provides a mapping from communication error types to repair strategies. When an automated system is able to recognize a particular error type, it can utilize the repair strategies corresponding to that type as identified in the research literature across disciplines. While several researchers have argued that systems need to be equipped with skills for rebounding when communication breaks down or other competencies are lacking (e.g., "strategic competence" for Canale ([5]), "miscommunication competence" for Perlis et al. ([22])), error recovery is abstracted as a type of meta-dialog competence along pragmatic dimensions. Here, the goal is to create a mapping from specific types of grounding problems to specific repair strategies. As a testament to the value of this kind of mapping, the next section highlights repair strategies corresponding to error types {0,1,1,1} and {0,0,1,1} that offer unique insights into effective error recovery.

3. Error recovery

In providing a mapping from communication error types to repair strategies, the classification scheme represents a valuable resource for system designer who want to exploit repair strategies that have been proven to be effective in other disciplines. In this section, we explore repair strategies for {0,1,1,1} and {0,0,1,1} that have been investigated by researchers outside of computer science. In particular, we discuss error recovery by people who experience some measure of language deficiencies as non-native speakers, hearing impaired, or aphasics.

3.1. Non-native speakers

In a study of how non-native speakers (NNS) deal with communication errors in conversation with native speakers (NS), Ulichny ([30]) found that repairs are extremely common; so much so that they "constitute the basic pattern of conversation between the speakers" (p.233). Corresponding to {0,0,1,1} in the classification scheme, the "trouble spot" for communication errors centered around a "mismatch of beliefs" concerning "the purpose of what was being said" (p.235).

The primary strategy a NNS uses as a listener to successfully handle possible communication errors is to "locally manage the misunderstanding;" that is, to resolve any uncertainties with clarification questions "within the same or immediately subsequent two conversational turns." In fact, in those instances in which a "conversational breakdown" occurred, defined as "areas of severe trouble in communicating speaker meaning leading to a disruption in the ongoing conversation" (p.234), the NNS postponed clarification only to reintroduce it out of sequence such as in the following example:

- NS1: uhm do you have any questions you want to ask *about the trip
- NNS: *oh yes about my level about my level=
- NS1: =well I *wasn't think about your level so much as about the=
- NNS: *ah because
- NS1: =trip –hh about your level in what sense
- NS2: I'll talk about the level afterwards when you're through
- NNS: ah . when the travel is uhm when uhm should I go to this
- NS1: when would be convenient for you
- NNS: when?
- NS1: when would be convenient for you
- NNS: ah . I don't know *perhaps in summer during summer
- NS1: *uh uhm yes that would be all right
- NNS: hmn no but I mm I mm wanted I I'd like to ask about my level for in CILTA for this
- NS1: that has nothing to do with me
- NS2: okay I'll talk to …

At the beginning of this interaction, the NNS makes an inquiry about his or her level. When the NS defers the question to another NS and moves on, the NNS seemingly accepts the response. However, the NNS eventually decides to come back to the issue for clarification, which leads to a disruption of the flow of conversation. Although the NNS in this instance probably decided to postpone clarification in order to maintain face, Ulichny point out that most attempts to return to a previous point of confusion ultimately have the opposite effect, causing the NS to retract positive judgment of his or her competence and to even engage in "foreigner talk – reducing the personality of the NNS to the level the NS assume for him" (p.244). The lesson for automated systems seems to be that in short, it is not cost effective to postpone clarification, and immediate, local management of uncertainty is critical.

As far as communication errors that a NNS encounters as a speaker, an extensive body of literature exists on "compensatory strategies" ([23], [28]) commonly utilized by a NNS when experiencing language production difficulties such as the use of elongated vowels and fillers to buy time for formulation.

3.2. Hearing impaired

Studies of the hearing impaired are particularly pertinent to system designers since communication errors with automated systems typically revolve around poor speech recognition. In one study, Tye-Murray & Witt ([29]) had adult cochlear-implant (CI) users interact in spontaneous, face-to-face conversation with normal, naïve partners who were unaware of their implants. Defining a "communication breakdown" as "instances in which persons with hearing impairment do not recognize spoken messages" (p.11), or {0,1,1,1} in the classification scheme, they found that CI users required 1.1 repair attempts to recover from communication breakdowns. This is consistent with the strategy of the NNS to locally manage uncertainties. Similar to the performance of a speech recognition engine, they also found that in 86% of communication breakdowns, CI users were aware that an utterance was spoken but did not recognize it. Otherwise, they were unaware, presumably due to error type {1,1,1,1}.

The repair strategies employed by CI users to minimize breakdowns included: taking longer speaking turns, interrupting more, using fillers, avoiding the use of questions, and shifting the topic (p.13). In applying these strategies to task oriented dialogs for automated systems, while all but the last two strategies may seem appropriate, other researchers have pointed out that the first two strategies do not facilitate pleasant conversations, though they may be effective. As Kaplan et al. ([17]) describe in their conversation workbook for the hearing impaired, three types of behaviors tend to emerge in response to communication difficulties: passive withdrawal from conversation through bluffing or pretending to understand, aggressive domination of conversation through interruptions, and assertive expression of communication needs. Recommending the last approach, Kaplan et al. suggest: "In order to use strategies successfully, a person must be assertive. That means the person must:

- 1. be willing to admit a hearing problem,
- 2. be willing to explain the problem to people when appropriate;
- 3. be able to suggest ways to improve communication" (p.18).

Among the repair strategies that were effective in recovering from breakdowns, Tye-Murray & Witt note that the top two strategies were non-verbal displays of confusion (38%) and confirmation questions (31%).

In summary, the lesson to be learned for system designer is again local management of uncertainties, using fillers to buy time if necessary, and taking an assertive approach to admitting misrecognitions. One spoken dialog system that has taken the assertive approach to heart is DeepListener ([16]), which engages in troubleshooting when misrecognitions are frequent. DeepListener even recommends that users try alternate commands to express their intention that are known to have better recognition accuracy; for instance, "I thought you might be saying 'Yes.' Just so you know I hear 'Sure' the best" ([16]).

3.3. Aphasics

In studies of how aphasics suffering from damage to the Broca and Wernicke areas of the brain engage in conversation with non-aphasics, researchers have found that aphasics regularly violate Gricean maxims of quantity and manner ([12]) and the principles of relevance, competence, and adequacy, as specified by Allwood ([2]). Yet they generally manage to keep the interaction going, despite the fact that they know they are producing "jargon" for their listeners at the {0,1,1,1} level.

Knowing that their listeners are struggling to recognize their speech, aphasics rely on a number of strategies to keep their conversation partners engaged. Their predominant strategy is to use of verbal and nonverbal politeness "to signal an attitude of cooperation and consideration" (p.63). This includes backchannels, laughter, and explicitly apologizing "for not being able to answer a question adequately" (p.63). In other words, aphasics exploit the joint activity nature of conversation by using politeness to remind listeners that coordination is required for activity to be successful.

System designers can make faster inroads into the handling of communication errors by construing dialog as a joint activity in which users are reminded that coordination and cooperation is required to successfully accomplish the joint task. When misrecognitions threaten to disrupt the task, it is useful to signal to users that the system is trying to keep up their part of the collaboration through backchannels, apologies, and "sincere" attempts at clarification.

4. Conclusion

In this paper, we offer to the community of researchers across disciplines with interest in communication errors, a classification scheme and taxonomy capturing how different types of communication errors are related to each other through the grounding of uncertainties. We also discussed how the scheme can be utilized for error handling in automated systems, and sketched an interdisciplinary taxonomy of communication error types. Finally, we examined repair strategies for error recovery used by people who, similar to automated systems, experience some measure of language deficiencies.

The classification scheme is presented with the hope that researchers across disciplines could more efficiently share results and findings about how to effectively deal with communication errors using grounding as a unifying theoretical framework.

5. Acknowledgements

The author would like to thank Eric Horvitz for insightful comments on previous drafts, and Herb Clark for discussions surrounding the value of the classification scheme.

6. References

[1] Ahlsen, E. 1993. Conversational principles and aphasic communication. *Journal of Pragmatics* 19(1): 57-70.

[2] Allwood, J. 1976. *Linguistic Communication as Action and Cooperation*. Gothenburg Monographs in Linguistics 2.

[3] Ardissono, L., Boella, G. & Damiano, R. 1998. A plan-based model of misunderstandings in cooperative dialogue. *Int. J. Human-Computer Studies* 48: 649-679.

[4] Blum-Kulka, S. & Weizman, E. 1988. The inevitability of misunderstandings: Discourse ambiguities. *Text* 8(3): 219-241.

[5] Canale, M. 1983. From communicative competence to communicative language pedagogy. In Richards, J.C., and Schmidt, R.W., eds., *Language and communication*. Longman.

[6] Clark, H.H. 1996. *Using Language*. Cambridge University Press.

[7] Clark, H.H. & Schaefer, E.F. 1987. Collaborating on contributions to conversations. *Language and Cognitive Processes* 2(1): 19-41.

[8] Clark, H.H. & Schaefer, E.F. 1989. Contributing to discourse. *Cognitive Science* 13: 259-294.

[9] Cohen, P.R. and Levesque, H.J.. 1994. Preliminaries to a collaborative model of dialogue. *Speech Communication* 15: 265-274.

[10] Danieli, M. 1996. On the use of expectations for detecting and repairing human-machine miscommunication. *AAAI-96*: 87-93

[11] Deemter, K. & Peters, S. 1996. *Semantic Ambiguity and Underspecification*. CSLI Publications.

[12] Grice, H. P. 1975. Logic and conversation. In Cole, P., and Morgan, J.L., eds., *Syntax and semantics*. Academic Press.

[13] Grosz, B.J. & Sidner, C.L. 1990. Plans for discourse. In Cohen, P., Morgan, J., and Pollack, M., eds., *Intentions in communication*. MIT Press.

[14] Gumperz, J. 1995. Mutual inferencing in conversation. In Markova, I, Graumann, C.F., and Foppa, K., eds., *Mutualities in dialogue*. Cambridge University Press.

[15] Hirst, G., McRoy, S., Heeman, P., Edmonds, P. & Horton, D. 1994. Repairing conversational misunderstandings and non-understandings. *Speech Communication* 15: 213-229.

[16] Horvitz, E. & Paek, T. 2001. Harnessing models of users' goals to mediate clarification dialog in spoken language systems. *User Modeling 2001*: 3-13.

[17] Kaplan, H., Bally, S. & Garretson, C. 1985. *Speechreading: A Way to Improve Understanding*. Gallaudet University Press.

[18] Marcu, D. & Hirst, G. 1996. A formal and computational characterization of pragmatic infelicities. *ECAI 1996*: 587-591

[19] McCoy, K.F. 1989. Generating context-sensitive responses to object misconceptions. *Artificial Intelligence* 41(2): 157-195.

[20] McRoy, S. & Hirst, G. 1995. The repair of speech act misunderstandings by abductive inference. *Computational Linguistics* 21(4): 435-478.

[21] Paek, T. & Horvitz, E. 2000. Conversation as action under uncertainty. *UAI 2000*: 455-464.

[22] Perlis, D., Purang, K. & Andersen, C. 1998. Conversational adequacy: mistakes are the essence. *Int. J. Human-Computer Studies* 48: 553-575.

[23] Poulisse, N., Bongaerts, T. & Kellerman, E. 1990. *The Use of Compensatory Strategies by Dutch Learners of English*. Foris Publications.

[24] Ringle, M. & Bruce, B. 1981. Conversation failure. In Lehnart, W. and Ringle, M., eds., *Strategies for natural language processing*. Lawrence Erlbaum Associates.

[25] Schegloff, E. 1987. Some sources of misunderstanding in talk-in-interaction. *Linguistics* 25(1): 201-218.

[26] Schegloff, E.A. & Sacks, H.. 1973. Opening up closings. *Semiotica* 8:289-327.

[27] Suchman, L.A. 1987. *Plans and Situated Actions: The Problem of Human/Machine Communication*. Cambridge University Press.

[28] Tarone, E. 1981. Some thoughts on the notion of communication strategy. *TESOL Quarterly* 15(3): 285-295.

[29] Tye-Murray, N. & Witt, S. 1996. Conversational moves and conversation styles of adult cochlear-implant users. *Journal of the Academy of Rehabilitation Audiology* 29: 11-25.

[30] Ulichny, P. 1997. The mismanagement of misunderstandings in cross-cultural interactions. *Journal of Pragmatics* 27(2): 233-246.