Using language

HERBERT H. CLARK | *Department of Psychology, Stanford University*

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011–4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 1996
Sixth printing 2005

Printed in the United Kingdom at the University Press, Cambridge

Set in Monotype Imprint and Monotype Grotesque

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data applied for

ISBN 0 521 56158 2 hardback
ISBN 0 521 56745 9 paperback

TAG

Preface

Writing a book can be like visiting a famous old city. You arrive with a copy of the *Guide Michelin* and begin touring the recommended sights. But as you walk from one landmark to the next, you discover the city beyond the *Guide*. Some features don't have the beauty or authenticity described in the *Guide*, and others aren't in the *Guide* at all. In one district, you find an exciting new style of architecture, and in another, an experiment in urban ecology. In still another, you come upon a new community of immigrants, complete with its own markets, restaurants, and religious activities. As you go from place to place, you meet more and more residents, who seduce you into extending your stay. By the time you leave, you realize that the city is just not what you expected. It is richer, more sophisticated, more diverse, and it took your visit to discover that.

Writing this book has been just such an experience. I am indebted to many for making it such an exciting, constructive, pleasurable, and prolonged experience. I wish to thank a great many collaborators for guiding me through new areas and expanding my horizons: Bridget Bly, Susan Brennan, Sam Buttrick, Stuart Card, Thomas Carlson, Jean Fox Tree, Ellen Francik, Wade French, Richard Gerrig, Ellen Isaacs, Barbara Malt, Catherine Marshall, Daniel Morrow, Gregory Murphy, Gisela Redeker, Edward Schaefer, Michael Schober, Robert Schreuder, Elizabeth Shriberg, Dale Schunk, Vicki Smith, Heather Stark, Elizabeth Wade, Thomas Wasow, Steve Whittaker, Deanna Wilkes-Gibbs. I owe a special debt to Randi Engle, Pim Levelt, Gisela Redeker, and Michael Schober for commenting on an earlier draft of the book and instigating fundamental changes in it. I credit Michael Schober with implanting the ideas that delayed the book the longest. Finally, the book wouldn't be what it is without Eve Clark, who has been the ideal companion on all my travels.

For the preparation of this book, I am indebted financially to the National Science Foundation, the Advanced Research Projects Agency, the Center for the Study of Language and Information at Stanford University, and, especially, the Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands.

Note on examples

A book about language use wouldn't be comprehensible without examples of spontaneous speech, so I have appealed to authentic examples wherever I could. Most of them are from the London–Lund corpus, a corpus of British English conversation collected and transcribed by Jan Svartvik, Randolph Quirk, and the Survey of English Usage at University College London and the Survey of Spoken English at the University of Lund (Svartvik and Quirk, 1980).[1] I have identified these examples by their text numbers (e.g., 1.1) and tone unit numbers (e.g., 245) like this: (1.1.245). The original transcripts represent tone units, intonation, overlapping speech, pauses, and many other features of spontaneous conversation. For readability, I have retained only some of these features, as illustrated here (1.1.245):

Reynard: so it's not until - next year that *the job will be advertised,*
Sam: *January I suppose there* may be an interview round about January,
Reynard: yeah, - u:m you heard anything about this, .
Sam: nothing at all yet, - -

This example contains the five special symbols:

Feature	Symbol	Example
End of tone unit	,	yeah,
Brief pause (of one light foot)	.	about this, . nothing
Unit pause (of one stress unit)	-	until - next year
Overlapping speech	*x* *y*	*the job will be advertised* *January I suppose there*
Elongated vowel	:	u:m

Overlapping speech, for example, is represented by two stretches of text enclosed by pairs of asterisks. Sam's "January I suppose there" overlaps with Reynard's "the job will be advertised." When there might be confusion, overlapping speech is enclosed in double asterisks, as in "**yeah**". Speech that was inaudible, or almost inaudible, to the transcriber is enclosed in double parentheses, as in "((3 or 4 sylls.))" or "((where are you))". Other noises are enclosed in single parentheses, as in "(- snorts)". In examples cited from other investigators, I have retained

[1] For analyses based on this corpus, see Erman (1987), Garnham, Shillcock, Brown, Mill, and Cutler (1982), Geluykens (1992), Oreström (1983), Stenström (1984), and Svartvik (1980).

the original notation, though sometimes in simplified form. On occasion I have highlighted the features of interest in boldface.

It is impossible to write about using language without mentioning the users themselves. In life, these users aren't generic speakers and addressees, but real people, with identities, genders, histories, personalities, and names. I have tried to keep this point in the foreground by giving the people in my examples names – their actual names whenever possible and fictitious names otherwise. The names serve to remind us of the subject matter of this book – that language is used by individuals at particular times and places for particular purposes.

PART I

Introduction

1 | Language use

Language is used for doing things. People use it in everyday conversation for transacting business, planning meals and vacations, debating politics, gossiping. Teachers use it for instructing students, preachers for preaching to parishioners, and comedians for amusing audiences. Lawyers, judges, juries, and witnesses use it in carrying out trials, diplomats in negotiating treaties, and actors in performing Shakespeare. Novelists, reporters, and scientists rely on the written word to entertain, inform, and persuade. All these are instances of *language use* – activities in which people do things with language. And language use is what this book is about.

The thesis of the book is this: Language use is really a form of *joint action*. A joint action is one that is carried out by an ensemble of people acting in coordination with each other. As simple examples, think of two people waltzing, paddling a canoe, playing a piano duet, or making love. When Fred Astaire and Ginger Rogers waltz, they each move around the ballroom in a special way. But waltzing is different from the sum of their individual actions – imagine Astaire and Rogers doing the same steps but in separate rooms or at separate times. Waltzing is the joint action that emerges as Astaire and Rogers do their individual steps in coordination, as a couple. Doing things with language is likewise different from the sum of a speaker speaking and a listener listening. It is the joint action that emerges when speakers and listeners – or writers and readers – perform their individual actions in coordination, as ensembles.

Language use, therefore, embodies both individual and social processes. Speakers and listeners, writers and readers, must carry out actions as individuals if they are to succeed in their use of language. But they must also work together as participants in the social units I have called ensembles. Astaire and Rogers perform both individual actions, moving their bodies, arms, and legs, and joint actions, coordinating these movements, as they create the waltz. In some quarters, language use has been studied as if it were entirely an individual process, as if it lay wholly within the cognitive sciences – cognitive psychology, linguistics, computer science, philosophy. In other quarters, it has been studied as if it were entirely a social process, as if it lay wholly within the social sciences – social psychology, sociology, sociolinguistics, anthropology. I suggest that it belongs to both. We cannot hope to understand language use without viewing it as joint actions built on individual actions. The challenge is to explain how all these actions work.

The goal of this chapter is to make a preliminary case for the thesis. To do this, I will take a tour through the settings of language use, the people who play roles in these settings, and the way joint actions emerge from individual actions. It will take the rest of the book to fill out the picture and develop principles to account for how language use is a joint action.

Settings of language use

Over the years, when I have asked people for instances of language use, they have offered such examples as "conversation," "reading a novel," "policemen interrogating a suspect," "putting on a play," "talking to oneself," and dozens more. These answers are remarkable for their range. To get a sense of that range, let us look at the answers classified by scene and medium. The scene is where the language use takes place.[1] The medium is whether the language use is spoken or signed or gestural, or written or printed, or mixed. I will use *setting* for the scene and medium combined and divide the media simply into *spoken* and *written* forms.

SPOKEN SETTINGS

The spoken setting mentioned most often is conversation – either face-to-face or on the telephone. Conversations may be devoted to gossip, business transactions, or scientific matters, but they are all characterized by the free exchange of turns among the two or more participants. I will call these *personal settings*. In monologues, in contrast, one person speaks with little or no opportunity for interruption or turns by members of the audience. Monologues come in many varieties

[1] See Hymes (1974, pp. 55-56) for a related use of setting and scene.

too, as when a professor lectures to a class, a preacher gives a sermon, or a student relates a recent experience to an entire class. These people speak for themselves, uttering words they formulated themselves for the audience before them, and the audience isn't expected to interrupt. These I will call *nonpersonal settings*.

In *institutional settings*, the participants engage in speech exchanges that resemble ordinary conversation, but are limited by institutional rules. As examples, think of a politician holding a news conference, a lawyer interrogating a witness in court, a mayor chairing a city council meeting, or a professor directing a seminar discussion. In these settings, what is said is more or less spontaneous even though turns at speaking are allocated by a leader, or are restricted in other ways. In *prescriptive settings*, in contrast, there may be exchanges, but the words actually spoken are completely, or largely, fixed beforehand. Think of the members of a church or synagogue reciting responsive readings from a prayer book, or a bride and groom reciting vows in a marriage ceremony, or a basketball referee calling foul. Prescriptive settings can be viewed as a subset of institutional settings.

The person speaking isn't always the one whose intentions are being expressed. The clearest examples are in *fictional settings*: John Gielgud plays Hamlet in a performance of *Hamlet*; Vivien Leigh plays Scarlett O'Hara in *Gone with the Wind*; Frank Sinatra sings a love song in front of a live audience; Paul Robeson sings the title role in the opera *Otello*; or a television pitchman makes a sales pitch to a television audience. The speakers are each vocalizing words prepared by someone else – Shakespeare, Cole Porter, the news department – and are openly pretending to be speakers expressing intentions that aren't necessarily their own.

Related to fictional settings are the *mediated settings* in which there are intermediaries between the person whose intentions are being expressed and the target of those intentions. I dictate a letter for Ed to my secretary Annie; a telephone company recording tells me of the time or weather; a television news reader reads the evening news; a lawyer reads Baker's last will and testament at a hearing; a recording is triggered in a building announcing a fire and describing how to find the fire escape; and a UN interpreter translates a diplomat's French simultaneously into English. When I dictate a letter to my secretary Annie and say "I'll see you Saturday," the person I expect to see on Saturday isn't Annie but the addressee of my letter Ed.

Finally, there are *private settings* in which people speak for themselves without actually addressing anyone else. I might exclaim silently to myself, or talk to myself about solving a mathematics problem, or rehearse what I am about to say in a seminar, or curse at another driver who cannot hear me. What I say isn't intended to be recognized by other people – at least in the way other forms of speaking are.[2] It is only of use to myself.

WRITTEN SETTINGS

When printing, writing, and literacy were introduced, people adapted spoken language to the printed medium, so it is no surprise that written uses have many of the characteristics of spoken ones. The written settings most like conversations are the personal settings, when people write to others they are personally acquainted with, as when I write my sister a letter, or write a colleague a message on the computer. In computer settings where the writing and reading on two terminals are simultaneous, the experience can resemble conversation even more closely.

Many written messages, however, are directed not at individuals known to the writer, but at a type of individual, such as "the reader of the *New York Times*" or "the reader of *Science*." These are *nonpersonal settings*. So a newspaper reporter writes a news story for readers of the *New York Times*, or an essayist writes on Scottish castles for readers of *Country Life*, or a physicist writes a textbook on electricity and magnetism for university undergraduates, or a car owner writes to the service department of Ford Motor Company. The reporter may know a few of the *New York Times'* readers, yet he or she is directing the news story at its general readership. Fiction, too, is usually directed at types of individuals, often defined very broadly, as when Henry James wrote *The Turn of the Screw*, and Edgar Allan Poe wrote "The Masque of the Red Death," and William Shakespeare wrote *Hamlet*. In written fiction, the author is writing for an audience, but as with spoken fiction, the intentions expressed are not his own.

Written settings, like spoken ones, can introduce intermediaries between the person whose intentions are being expressed and the intended audience. These again are mediated settings. Usually, the person actually writing the words is doing so in place of the person who appears to be doing the writing or speaking. Examples: The Brothers Grimm

[2] See the discussion of "response cries" (Goffman, 1978) in Chapter 11.

write down the folktale "Aschenputtel"; a translator translates *Hamlet* into French; a ghost writer writes Charlie Chaplin's autobiography; a speech writer writes a speech for the President; my secretary types the letter to Ed from my dictation; and the manuscript editor for this book edits my writing. The President's speech writers, for example, write as if they were the President, who later reads the words as if they were his or her own. We make the pretense that the speech writers weren't even involved in the process. Recorders, translators, ghost writers, secretaries, and manuscript editors, in their different ways, do much the same thing.

In some written settings, the words are selected through an institutional procedure. An advertising firm composes an advertisement for a magazine; a drug company composes the warning label for an aspirin bottle; a food company labels a package as baking soda; the US Senate legislates the wording of a new tax law; and the California legislature decides on the wording of state road signs. Although one person may have composed the words, it is the institution – the ad agency, drug company, or legislature – that is ultimately responsible, approving the wording as faithful to the institution's collective intentions.

Written language is used in private settings as well. I can write in my diary, scribble a reminder to myself, take notes on a lecture, make a grocery list, or work out a mathematics proof on paper. As in the spoken settings, I am writing solely to myself for later use.

What follows are examples of the major types of spoken and written settings, but these types are hardly exhaustive. Humans are creative. For each new technology – writing systems, printing, telegraph, telephones, radio, audio recording, television, video recording, telephone answering machines, interactive computers, and voice recognizers – people have developed new settings. With no end to new technologies, there is no end to the settings they might create. Our interest must be in the principles by which these new forms are created.

	Spoken settings	**Written settings**
Personal	A converses face to face with B	A writes letter to B
Nonpersonal	Professor A lectures to students in class B	Reporter A writes news article for readership B
Institutional	Lawyer A interrogates witness B in court	Manager A writes business correspondence to client B
Prescriptive	Groom A makes ritual promise to bride B in front of witnesses	A signs official forms for B in front of a notary public
Fictional	A performs a play for audience B	Novelist A writes novel for readership B
Mediated	C simultaneously translates for B what A says to B	C ghostwrites a book by A for audience B
Private	A talks to self about plans	A writes note to self about plans

CONVERSATION AS BASIC SETTING

Not all settings are equal. As Charles Fillmore (1981) put it, "the language of face-to-face conversation is the basic and primary use of language, all others being best described in terms of their manner of deviation from that base" (p. 152). If so, the principles of language use may divide mainly into two kinds – those for face-to-face conversation, and those that say how the secondary uses are derived from, or depend on it, or have evolved from it. Language uses are like a theme and variations in music. We look first at the theme, its melody, rhythm, and dynamics, and then try to discover how the variations are derived from it. Fillmore added, "I assume that this position is neither particularly controversial nor in need of explanation." Still, it is worth bringing out what makes face-to-face conversation basic and other settings not.

For a language setting to be basic, it should be universal to human societies. That eliminates written settings, since entire societies, and groups within literate societies, rely solely on the spoken word. One estimate is that about a sixth of the world's people are illiterate. And most languages as we know them evolved before the spread of literacy. We can also eliminate spoken settings that depend on such technologies as radio, telephones, television, and recordings, since these are hardly universal. Most people participate only rarely in nonpersonal, institutional, and prescriptive settings, and even then their participation is usually restricted to certain roles – audiences of lectures, parishioners, court observers. People do often participate in fictional settings, but usually as audience. The commonest setting is face-to-face conversation.

Face-to-face conversation, moreover, is the principal setting that doesn't require special skills. Reading and writing take years of schooling, and many people never do get very good at them. Even among people who know how to write, the most that many ever do is personal letters. Simple essays, to say nothing of news stories, plays, or novels, are beyond them. It also takes instruction to learn how to act, sing, lead seminars, chair meetings, and interrogate witnesses. And most people find it difficult to lecture, tell jokes, or narrate reasonable stories without practice. Almost the only setting that needs no specialized training is talking face to face.

Face-to-face conversation is also the basic setting for children's acquisition of their first language. For the first two or three years, children in both literate and illiterate societies learn language almost solely in conversational settings. Whatever they learn from books also comes in conversational settings, as their caretakers read aloud and check on what they understand. Children may learn some language from other media, but they apparently cannot learn their first language from radio or television alone.[3] In school, the language of peers is influential in the dialect acquired, and that too comes from conversational settings. Face-to-face conversation is the cradle of language use.

NONBASIC SETTINGS

What, then, makes other settings not basic? Let us start with the features of face-to-face conversation listed here (Clark and Brennan, 1991):

1	Copresence	The participants share the same physical environment.
2	Visibility	The participants can see each other.
3	Audibility	The participants can hear each other.
4	Instantaneity	The participants perceive each other's actions at no perceptible delay.
5	Evanescence	The medium is evanescent – it fades quickly.
6	Recordlessness	The participants' actions leave no record or artifact.
7	Simultaneity	The participants can produce and receive at once and simultaneously.

[3] For evidence, see Sachs, Bard, and Johnson (1981) and Snow, Arlman-Rupp, Hassing, Jobse, Joosten, and Vorster (1976).

8	Extemporaneity	The participants formulate and execute their actions extemporaneously, in real time.
9	Self-determination	The participants determine for themselves what actions to take when.
10	Self-expression	The participants take actions as themselves.

If face-to-face settings are basic, people should have to apply special skills or procedures whenever any of these features are missing. The more features are missing, the more specialized the skills and procedures. That is borne out informally.

Features 1 through 4 reflect the *immediacy* of face-to-face conversation. In that setting, the participants can see and hear each other and their surroundings without interference. Telephones take away copresence and visibility, limiting and altering language use in certain ways. Conversations over video hookups lack copresence, making them different too. In lectures and other nonpersonal settings, speakers have restricted access to their addressees, and vice versa, changing how both parties proceed. In written settings, which lack all four features, language use works still differently.

Features 5 through 7 reflect the *medium*. Speech, gestures, and eye gaze are evanescent, but writing isn't, and that has far-reaching effects on the course of language use. Speech isn't ordinarily recorded, but when it is, as on a telephone answering machine, the participants proceed very differently. In contrast, writing is ordinarily relayed by means of a printed record, and that leads to dramatic differences in the way language gets used. With written records and no instantaneity, writers can revise what they write before sending it off, and readers can reread, review, and cite what they have read. Most spoken settings allow the participants to produce and receive simultaneously, but most written settings do not. Being able to speak and listen simultaneously gives people in conversation such useful strategies as interrupting, overlapping their speech, and responding "uh huh," and these are ruled out in most written settings.

Features 8 through 10 have to do with *control* – who controls what gets done and how. In face-to-face conversation, the participants are in full control. They speak for themselves, jointly determine who says what when, and formulate their utterances as they go. In other settings, the participants are restricted in what they can say when. The church, for example, determines the wording of many prayers and responses. In fictional settings, speakers and writers only *make as if* they are taking certain actions – Gielgud is only play-acting his role as Hamlet – and that alters what they do and how they are understood. And in mediated settings, there are really two communications. Wim says "Heeft u honger?" in Dutch, which David translates for Susan as "Are you hungry?" Susan is expected to hear David's utterance knowing it is really Wim who is asking the question. The less control participants have over the formulation, timing, and meaning of their actions, the more specialized techniques they require.

What about private settings? These are sometimes considered the basic setting for language use. We all talk to ourselves, the argument goes, so private settings are surely universal. When we do talk to ourselves, however, the principal medium is the language we have acquired from others. People who know only English use English; people with only Chinese use Chinese; and people with only American Sign Language use American Sign Language. We may develop additional ways of talking to ourselves, but these too are derived from our social ways of talking. In talking to ourselves, we are making as if we were talking to someone else. Private settings are based on conversational settings.

In brief, face-to-face conversation is the basic setting for language use. It is universal, requires no special training, and is essential in acquiring one's first language. Other settings lack the immediacy, medium, or control of face-to-face conversation, so they require special techniques or practices. If we are ever to characterize language use in all its settings, the one setting that should take priority is face-to-face conversation. This is a point I will take for granted in the rest of the book.

Arenas of language use

Language settings are of interest only as arenas of language use – as places where people do things with language. At the center of these arenas are the roles of *speaker* and *addressee*. When Alan is addressing Barbara, he is the speaker and she the addressee. Now, Alan is speaking with the aim of getting Barbara to understand him and to act on that understanding. But he knows he cannot succeed unless she takes her own actions. She must attend to him, listen to his words, take note of his gestures, and try to understand what he means at the very moment he is speaking. Barbara knows all this herself. So Alan and Barbara don't act independently. Not only do they take actions *with respect to each other*, but they *coordinate* these actions with each other. In the term I introduced earlier, they perform joint actions. For a preview of how they manage that, let us start with the notion of background.

MEANING AND UNDERSTANDING

Alan and Barbara begin with a great mass of knowledge, beliefs, and suppositions they believe they share. This I will call their *common ground* (see Chapter 4). Their common ground may be vast. As members of the same cultural communities, they take as common ground such general beliefs as that objects fall when unsupported, that the world is divided into nations, that most cars run on gasoline, that *dog* can mean "canine animal," that Mozart was an eighteenth-century composer. They also take as common ground certain sights and sounds they have jointly experienced or that are accessible at the moment – gestures, facial expressions, and nearby happenings. And, finally, they assume to be common ground what has taken place in conversations they have jointly participated in, including the current conversation so far. The more time Alan and Barbara spend together, the larger their common ground.

Every social activity Alan and Barbara engage in takes place on this common ground (see Chapter 3). Shaking hands, smiling at one another, waltzing, and even walking past each other without bumping all require them to coordinate their actions, and they cannot coordinate their actions without rooting them in their common ground. When language is an essential part of the social activity, as it is in conversation or novel reading or play acting, there is an additional element of coordination between what speakers mean and what addressees understand them to mean – between *speaker's meaning* and *addressee's understanding.*

Suppose Alan points at a nearby sidewalk and says to Barbara "Did you see my dog run by here?" In taking these actions – his utterance, his gesture, his facial expression, his eye gaze – Alan means that Barbara is to say whether or not she saw his dog run by on the sidewalk he is pointing at. This special type of intention is what is called speaker's meaning (see Chapter 5). In doing what he did, Alan intends Barbara to recognize that he wants her to say whether or not she saw his dog run by on the sidewalk, and she is to see this in part by recognizing that intention. The remarkable thing about Alan's intentions is that they involve Barbara's thoughts about those very intentions. To succeed, he must get Barbara to coordinate with him on what he means and what she understands him to mean. That is a type of joint action.

Two essential parts of their joint action are Alan's signals and

Barbara's identification of those signals. I will use the term *signal* for any action by which one person means something for another person. That is, meaning and understanding are created around particular events – with qualifications to come later – that are initiated by speakers for addressees to identify. These events are signals. Alan's signal consists of his utterance, gestures, facial expression, eye gaze, and perhaps other actions, and Barbara identifies this composite in coming to understand what he means (see Chapter 6).

Signals are deliberate actions. Some are performed as parts of conventional languages like English, Dakota, Japanese, or American Sign Language, but any deliberate action can be a signal in the right circumstances. Juliet signaled Romeo that it was safe to visit by hanging a rope ladder from her window. Umpires and referees signal fouls and goals with conventional gestures. Good storytellers signal aspects of their descriptions with nonconventional depictive gestures. We all signal things with deliberate smiles, raised eyebrows, empathetic winces, and other facial gestures. We even signal things by deliberately failing to act where such an action is mutually expected – as with certain pauses and deadpan expressions.[4] So some aspects of signals are conventional, and others are not. Some of the conventional aspects belong to systems of signals such as English or American Sign Language, and others do not. And some signals are performed as parts of intricate sequences, as in conversation or novels, and others are not. When Juliet hung a ladder out for Romeo, she created an isolated signal for a special purpose.

It is impossible for Alan and Barbara to coordinate meaning and understanding without reference to their common ground. When Alan says, "Did you see my dog run by here?" Barbara is to consult the meanings of the words *did*, *you*, *see*, etc., and their composition in English sentence constructions. These meanings and constructions are part of Alan and Barbara's common ground because Alan and Barbara are both members of the community of English speakers. To recognize the referents of *my*, *you*, *here*, and the time denoted by *did see*, Barbara is to take note of other parts of Alan's signal – that he is gazing at her now and gesturing at a nearby sidewalk. That in turn requires her to consult their

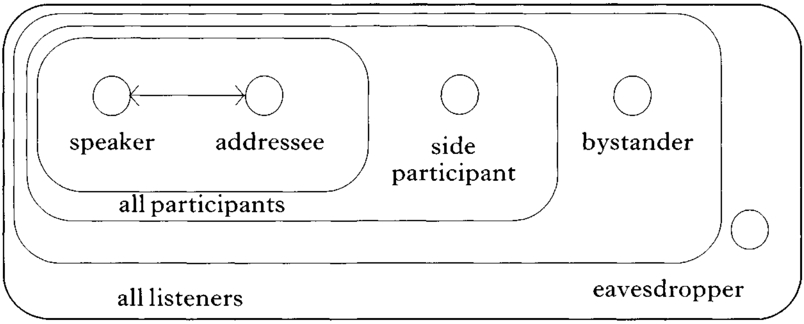
[4] A more accurate name for language use might be signal use, since it doesn't suggest an exclusive concern with conventional languages. Unfortunately, such a term is more likely to appeal to generals or engineers than to the rest of us. It would never catch on.

common ground about the immediate situation – that they are facing each other, that the sidewalk is nearby, that Alan is scanning the area in search of something. To identify the referent of *my dog*, she is to consult their common ground for a unique dog associated with him. Common ground is the foundation for all joint actions, and that makes it essential to the creation of speaker's meaning and addressee's understanding as well.

PARTICIPANTS

When Alan asks Barbara about his dog, Connie may also be taking part in the conversation, and Damon may be overhearing from nearby. Alan, Barbara, Connie, and Damon each bear a different relation to Alan's question.

The people around an action like Alan's divide first into those who are truly participating in it and those who are not: *participants* and *non-participants*. For Alan's question, the participants are Alan himself, Barbara, and Connie. These are the people he considers "ratified participants" (Goffman, 1976). They include the speaker and addressees – here Alan and Barbara – as well as others taking part in the conversation but not currently being addressed – here Connie. She is a *side participant*. All other listeners are *overhearers*, who have no rights or responsibilities in it. Overhearers come in two main types. *Bystanders* are those who are openly present but not part of the conversation. *Eavesdroppers* are those who listen in without the speaker's awareness. There are in reality several varieties of overhearers in between.



Alan must pay close attention to these distinctions in saying what he says. For one thing, he must distinguish addressees from side participants. When he asks Barbara about his dog and Connie is in the conversation, he must make sure they see that it is Barbara, and not

Connie, who is to answer his question. Yet he must make sure Connie understands what he is asking Barbara (see Chapter 3). He must also take account of overhearers, but because they have no rights or responsibilities in the current conversation, he can treat them as he pleases. He might, for example, try to conceal from Damon what he is asking Barbara by saying "Did you happen to see you-know-what come by here?" It isn't always easy to deal with participants and overhearers at the same time (Clark and Carlson, 1982a; Clark and Schaefer; 1987a, 1992; Schober and Clark, 1989).

So side participants and overhearers help shape how speakers and addressees act toward each other. They also represent different ways of listening and understanding. As an addressee, Barbara can count on Alan having designed his utterance for her to understand, but as an overhearer, Damon cannot. As a result, the two of them go about trying to interpret what Alan says by different means, by different processes. These other roles should help us see more precisely what the roles of speaker and addressee themselves are, and they will.

LAYERS IN LANGUAGE ARENAS

The roles we have met so far, from speaker to eavesdropper, may each enter into a primary setting with a single place, time, and set of participants. In other settings, other agents may take part too, including authors, playwrights, mediators, actors, ghost writers, translators, and interpreters, and they may take part at different places and times. How are we to characterize these other places, times, and roles? What we need, I will suggest, is a notion of layering (Chapter 12).

When someone tells a joke, the other participants must recognize it for what it is – a piece of fiction. Take this stretch of conversation (from Sacks, 1974, in simplified format):

Ken: You wanna hear- My sister told me a story last night.
Roger: I don't wanna hear it. But if you must. (0.7)
Al: What's purple and an island. Grape, Britain. That's what his sister told him.
Ken: No. To stun me she says uh, (0.8)
There were these three girls and they just got married?
[Continues joke]

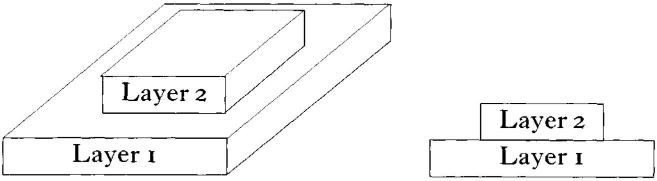
When Ken says "My sister told me a story last night," he is making an assertion to Roger and Al in the actual world of the conversation. But when he says "There were these three girls and they just got married," he is making an assertion that is true only in the hypothetical world of the joke. He doesn't really believe there were three actual girls who just got married. He is speaking at that moment as if he, Roger, and Ken were part of the hypothetical joke world, and he was telling them about three actual girls.

What we have here are two *layers of action*. Layer 1 is the primary layer of any conversation, where the participants speak and are addressed then and there as themselves. Layer 2 is built on top of layer 1 and in this example represents a hypothetical world. Each layer is specified by its domain or world – by who and what are in it. When Ken says "My sister told me a story last night," his actions take place entirely in layer 1, the actual domain of their conversation. But when he says "There were these three girls and they just got married," he is both making an assertion in layer 2, the hypothetical domain of the joke, and telling part of a joke in layer 1, the actual domain:

Layer 2 Ken is telling Roger and Al about three actual girls who just got married.
Layer 1 In Los Angeles in 1965, Ken, Roger, and Al jointly pretend that the events in layer 2 are taking place.

We would say that Roger and Al had misunderstood Ken if they thought that the sister was hypothetical and the three girls were actual. Language use requires the primary participants to recognize, however vaguely, all the layers present at each moment.

Layers are like theater stages built one on top of another. In my mind's eye, they look like this:



Layer 1 is at ground level, representing the actual world, which is present in all forms of language use. Layer 2 is a temporary stage built on top of layer 1 to represent a second domain. As on a theater stage, characters perform actions in full view of the participants of layer 1. As on a theater stage, these characters cannot know that layer 1 even exists. The three girls have no way of knowing about Ken, Roger, and Al's conversation. In this picture, layer 1 is real, whereas layer 2 is optional and only supported by layer 1. And by recursion there can be higher layers as well.

With layering we can now represent what makes many language settings derivative (see Chapter 12). Face-to-face conversation and personal letters are normally managed in one layer. Jokes, novels, and other pieces of fiction take at least two layers, and when a school teacher reads a piece of fiction aloud, that adds yet another layer. Plays require at least three layers. Dictation also requires two layers. When I dictate a letter for my friend to my secretary, I am talking to my secretary at layer 1 – our actual conversation – yet, simultaneously, speaking to my friend at layer 2. Ghost writing, simultaneous translation, and news reading require still other patterns of layering.

Layering also helps make sense of private uses of language. When George curses at a bad driver who cannot hear him, he deals in two layers. In the privacy of his car (layer 1), he creates in his imagination a domain (layer 2) in which he is actually cursing the driver face to face. When Helen silently exclaims to herself about a beautiful sunset, she does much the same thing. In private, layer 1, she creates an imaginary domain (layer 2) in which she is speaking to her alter ego. With diaries, reminders, and grocery lists, the writers are addressing themselves at a later time and place. This is no different from writing to someone else at a later time and place.

So far, we have seen that language use places people in many roles. In basic settings, there are always speakers and addressees, but there may also be side participants, bystanders, and eavesdroppers. In other settings, there may also be more than one layer of activity, each with its own roles. The primary layer, which I have called layer 1, represents actual people doing actual things. Higher layers represent other domains, often hypothetical, that are created only for the moment. It often takes many different roles, such as actor and stenographer, to create and support them.

Actions of language

What people do in arenas of language use is take actions.[5] At a high level of abstraction, they negotiate deals, gossip, get to know each other. At a lower level, they make assertions, requests, promises, apologies to each other. In doing that, they categorize things, refer to people, and locate

[5] By *action*, *act*, and *activity*, I shall always mean doing things *intentionally*. For two views of intention and action, see Bratman (1987, 1990) and Cohen and Levesque (1990).

objects for each other. At yet a lower level, they produce utterances for each other to identify. And at the lowest level, they produce sounds, gestures, writing for each other to attend to, hear, see. These at least are the actions of speakers and addressees in the primary layer of language use. Strikingly, all these actions appear to be joint actions – an ensemble of people doing things in coordination. If we are ever to understand them, we need to know what joint actions are and how they work. That is the topic of Chapter 3. For now, let us look briefly at joint actions and how they are created out of individual actions.

JOINT ACTIONS

When I play a Mozart sonata on the piano, the music I produce reflects certain of my mental and motor processes, from reading the printed music to striking the keys with my fingers. These processes are wholly under my control – as afforded by the piano's mechanics, the printed score, the lighting, and other environmental features. I decide when to begin, how fast to play, when to slow down or speed up, when to play forte and when pianissimo, and how to phrase things. And if my mental and motor processes come off just right, the result will be Mozart.

Something different happens when a friend, Michael, and I play a Mozart duet. This time, my actions depend on his, and his depend on mine. We have to coordinate our individual processes, from reading the notes to striking the keys. Each decision – when to begin, how fast to go, when to slow down or speed up, when to play forte and when pianissimo, how to phrase things – must be a joint one, or the result won't be Mozart. Our performance is best described not as *two individuals* each playing a Mozart piece, but as a *pair of people* playing a Mozart duet.

One contrast here is between *individual* and *joint* actions. A joint action is an action by an ensemble of people. Playing solo is an individual action, but playing a duet is a joint one. We see the same contrast in these comparisons:

Individual action	**Joint action**
A person paddling a kayak	A pair of people paddling a canoe
A person pushing a car	A quartet of people pushing a car
A lumberjack cutting a log with a saw	A pair of lumberjacks cutting a log with a two-handled saw
A ballerina dancing to a recording	A corps de ballet dancing to a recording
A race-car driver speeding around a track	A set of ten race-car drivers speeding around a track

A person's processes may be very different in individual and joint actions even when they appear identical. Suppose I play my part of the Mozart duet on an electronic keyboard twice – once solo and once with Michael playing his part. If you listened to my part through earphones, you might not notice any difference, yet what I did was very different. In the solo performance I took every action on my own. In the duet I coordinated every action with Michael, and as anyone who has played duets knows, that is no small feat. There are analogous differences between one and two canoe paddlers, one and four auto pushers, one and many dancers, one and two lumberjacks, and one and ten race-car drivers. All these cases illustrate the same point: Performing an individual action solo is not the same as performing the apparently identical action as part of a joint action.

We must therefore distinguish two types of individual actions. When I play the piano solo, I am performing an *autonomous action*. When Michael and I play the piano duet, we are also performing individual actions, but as parts of the duet. These actions are what I will call *participatory actions*: They are individual acts performed only as parts of joint actions. So joint actions such as playing piano duets are constituted from participatory actions. Or, what is the same thing, it takes participatory actions to create joint actions. They are two sides of the same coin:

Type of action	**Agents**
joint actions	ensemble of participants
participatory actions	individual participants

We can look at any joint action either way – as a whole made up of parts, or as parts making up the whole.

Many joint actions have the participants doing dissimilar things. A driver approaching a crosswalk coordinates with the pedestrian trying to cross it. A ballerina dancing coordinates with the orchestra accompanying her. A clerk slipping a shoe on a woman's foot coordinates with the woman as she extends her foot to accept it. These examples make a second point about joint actions: The participants often perform very different individual actions.

SPEAKING AND LISTENING

Speaking and listening have traditionally been viewed as autonomous actions, like playing a piano solo. One person, say Alan, selects and produces a sentence in speech or on paper, and another person, say

Barbara, receives and interprets it. Using language is then like transmitting telegraph messages. Alan has an idea, encodes it as a message in Morse code, Japanese, or English, and transmits it to Barbara. She receives the message, decodes it, and identifies the idea Alan wanted her to receive.[6] I will argue that speaking and listening are not independent of each other. Rather, they are participatory actions, like the parts of a duet, and the language use they create is a joint action, like the duet itself.

Speaking and listening are themselves composed of actions at several levels. As Erving Goffman (1981a, p. 226) noted, the commonsense notion of speaker subsumes three agents.[7] The *vocalizer* is "the sounding box from which utterances come." (The corresponding role in written settings might be called the *inscriber.*) The *formulator* is "the agent who puts together, composes, or scripts the lines that are uttered." And the *principal* is "the party to whose position, stand, and belief the words attest." The principal is the agent who *means* what is represented by the words, the *I* of the utterance. In Goffman's view, speaking decomposes into three levels of action: meaning, formulating, and vocalizing (see also Levelt, 1989).

In face-to-face conversations, the speaker plays all three roles at the same time – principal, formulator, and vocalizer. When Alan asks Barbara "Did you happen to see my dog run by here?" he selects the meaning he wants to be recognized; he formulates the words to be uttered; and he vocalizes those words. In nonbasic settings, these roles often get decoupled. When a spokeswoman reads a statement by the Secretary of State, she vocalizes the announcement, but it is the Secretary whose meaning she represents, and an aide who formulated them. Ghost writers, to take a different case, formulate and inscribe what they write, but their words represent the meanings of the people they are ghosting for. Much the same goes for translators, speech writers, and copy editors. And in prescriptive settings, meaning and vocalizing get decoupled from formulating. When a bride says "I Margaret take thee Kenneth to my wedded husband" in a marriage ceremony, she refers to herself with *I*,

[6] The *message model* implies that Alan's production, and Barbara's reception, can be studied in isolation. It also implies that messages are encoded strings of symbols in a symbol system (say, Japanese or English), so they can be studied in isolation from the processes by which they are produced and received. If speaking and listening are participatory actions, these two implications no longer follow.

[7] To avoid confusion, I have replaced Goffman's terms *animator* and *author* by the terms *vocalizer* and *formulator*.

meaning what she says, but she doesn't formulate what she says. That is prescribed by the church.

Listening, likewise, decomposes into at least three levels of action. When Barbara is asked by Alan "Did you happen to see my dog run by here?" she is first of all *attending* to his vocalizations. She is also *identifying* his words and phrases. And she is the *respondent*, the person who is to recognize what he meant and answer the question he asked. In face-to-face conversations, the addressee plays all three roles at once – respondent, identifier, and attender. But in nonbasic settings, once again, the roles often get decoupled. The main job of copyists, court reporters, and stenographers, for example, is to identify people's utterances, though it is typical for them to try to understand as they do that. Or when Wim, speaking Dutch, says something to Susan through a simultaneous translator speaking English, she may attend to Wim's utterances without identifying or understanding them. And although she attends to, identifies, and understands the translator's English, the only thing she attributes to Wim is the meaning expressed.

The component actions in speaking and listening come in pairs. For each action in speaking, there is a corresponding action in listening:

	Speaking	**Listening**
1	A vocalizes sounds for B	B attends to A's vocalizations
2	A formulates utterances for B	B identifies A's utterances
3	A means something for B	B understands A's meaning

But the pairing is even tighter than that. Each level consists of two participatory actions – one in speaking and one in listening – that together create a joint action. The overall joint action really decomposes into several levels of joint actions. This is a topic I take up in Chapters 5, 7, 8, and 9.

One of these joint actions is privileged, and it is level 3: speaker's meaning and addressee's understanding. It is privileged, I suggest, because it defines language use. It is the ultimate criterion we use in deciding whether something is or is not an instance of language use. Language use, I assume, is what John Stuart Mill called a *natural kind*.[8] It is a basic category of nature, just as cells, mammals, vision, and learning are, one that affords scientific study in its own right. And what makes it a natural kind is the joint action that creates a speaker's meaning and an addressee's understanding.

[8] See, for example, Quine (1970) and Putnam (1970).

EMERGENT PRODUCTS

When we take an action, we foresee, even intend, many of its consequences, but other consequences simply emerge. That is, actions have two broad products: *anticipated products* and *emergent products.* Let us consider some examples.

A friend tells you to print the words *slink, woman, ovate, regal,* and *droll* one below the other, and you do. Then she says, "Now read down the five columns," and you discover, to your amazement, five more words: *sword, lover, imago, natal,* and *knell* (from Augarde, 1986). The down words weren't anything you anticipated. They just emerged. Then you take your discovery to another friend. "Let me print the words *slink, woman, ovate, regal,* and *droll* one below the other. See the words that you get reading down." This time you intend to form the words reading down, so they become an anticipated product.

A twelve-year-old tells you, "Say E," and you say "E." "Say S," and you say "S." "Say X," and you say "X." "Say E," and you say "E." The child says "Now say them all, quickly, three times" and you say "ESXEESXEESXE." And the child retorts "No he isn't!" In producing "ESXE" quickly, you didn't anticipate it would sound as if you were saying "He is sexy." That was an emergent product of your action.

Susan composes a mystery duet for Michael and me to play on two pianos. Our parts are so cleverly devised that neither of us can tell what the duet will sound like. The day we perform it together we discover we are playing "Greensleeves." Later we go to other friends, announce that we are going to play "Greensleeves," and each play our parts. On the first performance, "Greensleeves" was an emergent product of our joint actions, but on the second, it is the anticipated, even intended, product.

When individuals act in proximity to each other, the emergent product of their actions may even go against their desires, a point made by Thomas Schelling (1978). Individuals enter an auditorium one by one. The first arrival sits one third of the way back – not too far forward, but not too far back either. The second and later arrivals, to be polite, choose to sit behind the front-most person. As the auditorium fills, the pattern that emerges has everyone in the rear two thirds of the auditorium. Each individual might prefer the audience to be in the front two thirds of the auditorium, but they have to live with the pattern that emerged.

All actions have anticipated products, and that goes for joint actions too. When Michael and I played our parts of the Mozart duet, we intended to produce the Mozart duet. It was anticipated. Joint actions also have emergent products. When Michael and I played Susan's duet for the first time, we intended to "play a duet," but we didn't intend to "play 'Greensleeves.'" It is simply what emerged. In language use, it is important not to confuse anticipated and emergent products. Many of the regularities that are assumed to be intended or anticipated are really neither, but simply emerge.

SIX PROPOSITIONS

In this chapter I have sketched the approach to language use I will take in this book. Along the way I have introduced several working assumptions.

Proposition 1. Language fundamentally is used for social purposes. People don't just use language. They use language for doing things – gossiping, getting to know each other, planning daily chores, transacting business, debating politics, teaching and learning, entertaining each other, holding trials in court, engaging in diplomacy, and so on. These are social activities, and language is an instrument for helping carry them out. Languages as we know them wouldn't exist if it weren't for the social activities they are instrumental in.

Proposition 2. Language use is a species of joint action. All language use requires a minimum of two agents. These agents may be real or imaginary, either individual people or institutions viewed as individuals. In using language, the agents do more than perform autonomous actions, like a pianist playing solo. They participate in joint actions, like jazz musicians improvising in an ensemble. Joint actions require the coordination of individual actions whether the participants are talking face to face or are writing to each other over vast stretches of time and space.

Proposition 3. Language use always involves speaker's meaning and addressee's understanding. When Alan produces a signal for Barbara to identify, he means something by it: He has certain intentions she is to recognize. In coordination with him, Barbara identifies the signal and understands what he means by it. Much of what we think of as language use deals with the mechanics of doing this effectively. We are not inclined to label actions as language use unless they involve one person meaning something for another person who is in a position to understand what the first person means. Proposition 3 doesn't imply, of course, that language use is nothing more than meaning and understanding. It is a great deal more. It is just that these notions are central, perhaps criterial, to language use.

Proposition 4. The basic setting for language use is face-to-face conversation. For most people conversation is the commonest setting of language use, and for many, it is the only setting. The world's languages have evolved almost entirely in spoken settings. Conversation is also the cradle for children learning their first language. It makes no sense to adopt an approach to language use that cannot account for face-to-face conversation, yet many theorists appear to have done just this. And if conversation is basic, then other settings are derivative in one respect or another.

Proposition 5. Language use often has more than one layer of activity. In many types of discourse – plays, story telling, dictating, television news reading – there is more than one domain of action. Each domain is specified by, among other things, a set of participants, a time, a place, and the actions taken. The actions that story tellers take toward their audience, for example, are in a different layer from the actions that the fictional narrators in their stories take toward their fictional audiences. Conversation, at its simplest, has only one layer of action. The speaker at any moment is the principal, formulator, and vocalizer of what gets said, and the addressees are attenders, identifiers, and respondents. Still, any participant can introduce further layers of action by telling stories or play-acting at being other people. This makes conversation one of the richest settings for language use.

Proposition 6. The study of language use is both a cognitive and a social science. We can view a joint activity such as playing a piano duet from two perspectives. We can focus on the individual pianists and the participatory actions they are each performing. Or we can focus on the pair and the joint action they create as a pair. For a complete picture, we must include both. We cannot discover the properties of playing duets without studying the pianists playing as a pair, and yet we cannot understand what each pianist is doing without recognizing that they are trying to create the duet through their individual actions.

Although the study of language use ought to resemble the study of any other joint activity, it doesn't. Cognitive scientists have tended to study speakers and listeners as individuals. Their theories are typically about the thoughts and actions of lone speakers or lone listeners. Social scientists, on the other hand, have tended to study language use primarily as a joint activity. Their focus has been on the ensemble of people using language to the neglect of the thoughts and actions of the individuals. If language use truly is a species of joint activity, it cannot be understood from either perspective alone. The study of language use must be both a cognitive and a social science.

In this book I combine the two views. In Part II, I take up three foundations of language use: the notion of broad joint activities (Chapter 2), the principles behind joint actions (Chapter 3), and the concept of common ground (Chapter 4). In Part III, I turn to communicative acts themselves, developing the notions of meaning and understanding (Chapter 5) and signaling (Chapter 6). In Part IV, I explicate the notion of levels in joint actions, arguing for a level of joint projects (Chapter 7), meaning and understanding (Chapter 8), presenting and identifying utterances, and executing and attending to behaviors (Chapter 9). In Part V, I take up three broader issues: the joint commitments established in exchanges of goods (Chapter 10); features of conversation (Chapter 11); varieties of layering (Chapter 12). In Part VI, I conclude.