Production–Comprehension Interface

Intermediate article

Victor S Ferreira, University of California, San Diego, California, USA

CONTENTS

Introduction
Inclusion of information based on listener knowledge
Modifications of the form of spoken utterances
Modifications of the content of spoken utterances
Common ground effects
The nature of common ground
Processing considerations

Many features of spoken language, from the nature of child-directed speech to audience design effects, reflect a sensitivity in language production to the needs and strategies of language comprehension. These adjustments, sometimes deliberate, sometimes automatic, ensure successful communication.

INTRODUCTION

The primary reason that speakers speak is so that their listeners can understand them. It is therefore unsurprising that unlike many cognitive tasks (such as perceiving, remembering, or decision-making), the language production performance of an individual speaker must take into account the processing capabilities and the knowledge states of another, namely, that speaker's intended listener. Research into this topic of the nature of the production–comprehension interface generally explores how production accommodates its processing to the needs of comprehension.

This article discusses the production–comprehension interface by exploring two related issues. The first concerns how speakers cater specific details of their utterances to take into account the knowledge and the comprehension capabilities and strategies of their listeners. The second narrows in on how speakers use common ground in their utterances – information that the speaker and listener believe to be mutually known in their present conversation or discourse.

INCLUSION OF INFORMATION BASED ON LISTENER KNOWLEDGE

The ways that speakers and listeners try to accommodate one another in conversations was described by Grice (1975). A Gricean approach to language specifies that interlocutors tacitly agree to the cooperative principle when they participate in a conversation, namely, that they agree to be mutually productive and efficient participants in their conversation. The cooperative principle is instantiated by four specific maxims. The maxim of *quality* states that interlocutors mutually assume that contributions to a conversation are legitimately provided as statements of truth. Thus, if a speaker says, 'it's raining', she or he has some evidentiary basis upon which to make that statement, and she or he does not intend the listener to understand that it is snowing. The maxims of *relation* and of *quantity* are relevant to the issue of common ground discussed below. The maxim of relation states that interlocutors are assumed to make their current utterances relevant by ensuring that current references relate to already known information. The maxim of quantity states that interlocutors are assumed to contribute an appropriate amount of information with each utterance, providing neither too little nor too much information (which cannot be done unless the speaker knows what the listener already knows). Finally, the maxim of *manner* states that interlocutors should endeavor to make their utterances unambiguous, brief, orderly, and that they should avoid obscurity. The maxim of manner can thus be taken to claim that speakers should make the form of their utterances easy for a listener to understand. Note that these maxims apply both to speakers and to listeners, in that speakers use the principle to guide utterance formation, and listeners use the principle to guide utterance interpretation. Furthermore, the cooperative principle can be flouted, or seemingly violated for the sake of communication (creating an implicature). For example, when asked how a tennis match was, if

a speaker says 'well, it was a tennis match', she or he seems to violate the maxim of quantity (by providing too little information), but thereby communicates that not much is to be said about the tennis match.

Grice's maxims are of two importantly distinct types. The maxims of quantity, quality, and relation are about what a speaker chooses to mention, whereas the maxim of manner is about how a speaker describes what they have decided to mention. This latter issue is discussed next, describing how speakers modify the form of their utterances to make those utterances easy to understand. Then, the issue of what a speaker chooses to mention is described, leading to the issue of the use of common ground.

MODIFICATIONS OF THE FORM OF SPOKEN UTTERANCES

Speakers cater the form of their utterances to their listeners in two general ways. One way involves explicit recognition by the speaker that their listener may encounter comprehension difficulty, typically because the listener may have limited linguistic abilities. Here, speakers are consciously aware of the potential for communicative difficulty, and they thereby make general modifications to their speech to circumvent that difficulty. The second way involves implicit production strategies that speakers use to create easy-to-understand sentences. Here, speakers are typically unaware of the difficulty their listener may have, but nevertheless tacitly make very specific modifications of their speech to enhance comprehensibility. Each of these kinds of modifications are discussed in turn.

When a speaker addresses a listener, the general style or register of his or her speech will be partially based on the nature of the relationship between the speaker and listener. Thus, for example, a speaker generally adopts a different register when addressing parents or grandparents, compared to when speaking to siblings. Students say 'yes, professor' to a university professor, but talk about 'my prof' to their friends.

While many of the features of a speech style reflect or communicate something about the social relationship between speaker and listener, other register effects may reflect a speaker's estimation of the linguistic capabilities of his or her listener. This is most strongly revealed by analyses of speech directed to populations with limited linguistic capabilities, including speech directed to children (child-directed speech or 'motherese'), to foreign-language speakers (for the purpose of communication or for language instruction), and to cognitively impaired listeners (e.g. the mentally retarded). When addressing populations like these, speakers (intentionally) include a host of features in their utterances that (intentionally or unintentionally) serve two related purposes: speakers modify their utterances so that they are more easily understood, and so that they are more likely to capture and hold their listener's attention. The features of utterances that accomplish these goals are listed in Table 1.

In addition to these general modifications that occur in response to perceived characteristics of a speaker's current listener, speakers make other, more specific modifications that occur automatically. Unlike the register effects just described, these more specific changes do not depend on the characteristics of a speaker's listener, and indeed occur in the absence of any listener at all.

For example, one bit of information that is useful for a listener to know is whether a mentioned entity has previously been mentioned in a discourse, or whether that entity is new to a discourse (this is part of the information structure of a discourse). Languages have a number of devices to mark this difference between given information and new information, including lexical, syntactic, and acoustic devices. Lexically, speakers can refer to given

Table 1. Features of a speech register that make it easier to understand and more likely to capture and maintain listeners' attention

Features that make speech easier to understand	*Features that make speech more attention-getting*
Shorter utterances	Greater use of names
More repetition of words and phrases	More questions
More rephrasals	Higher and more modulated pitch
More common words	Longer vowels
Simpler and more transparent syntactic structures	Longer pauses
Hyperarticulation	Greater rhythm

information with pronouns ('she', 'he', etc.) and by using definite articles ('the', 'this', etc.). Syntactically, speakers mark the difference between given and new information typically by word order, where speakers tend to mention given entities in their utterances prior to new entities. For example, if the sentence 'I saw Bill yesterday' sets 'Bill' up as given information, it is more natural to say 'he was disappointed by the election', with the given information first, compared to 'the election disappointed him', with the given information second. Finally, acoustically, speakers tend to diminish the auditory characteristics of words that refer to given information (affecting the prosody of those words), by reducing the duration, the loudness, and the pitch of those words. These cues to the given versus new status of information are used redundantly, in that the use of one kind of cue does not preclude the use of another (e.g. given information might be produced early in a sentence and in reduced fashion).

Speakers use other subtle signals to communicate useful information to their listeners. Some of these signals arise when speakers anticipate upcoming or recognize past production difficulty. For example, speakers use the familiar fillers 'uh' and 'um' differently, in that 'uh' is used when speakers anticipate a relatively short disruption to their own speech, whereas 'um' is used when speakers anticipate a longer disruption. Similarly, speakers tend to use 'thee' instead of 'the' when they are about to have difficulty retrieving the following word (Fox Tree and Clark, 1997). When speakers repair an erroneously produced part of an utterance, they typically say 'er' or 'I mean' to mark that a repair is to follow, and they reproduce enough of the original utterance so that the utterance plus repair are easily understood. Other constraints (termed *perceptual constraints*; see Cutler, 1987) include a tendency to speak with a consistent rhythm (which assists listeners in finding word boundaries), and a tendency to use easily recognizable forms when creating novel words (e.g. speakers prefer to say 'ambiguize' over 'ambigwify', because the former sounds more similar to the root form, 'ambiguous').

It is important to recognize that these specific modifications occur unintentionally and automatically. For example, the tendency to manipulate word order so that given information occurs before new information occurs in the complete absence of any listener, in a manner that appears to be sensitive to how easily retrieved an entity is from the speaker's memory. The tendency to diminish the acoustic characteristics of words referring to given information has been found when speakers address 14-month-old children (who, at this largely preconversational stage of linguistic development, are unlikely to usefully distinguish given from new information), and when speakers produce the first and the second mention of a word to different listeners. This is different from what occurs with the register effects discussed above, which are generally dependent upon the presence of a linguistically challenged interlocutor for the appropriate register features to appear correctly.

Different mechanisms are used by the language production system to automatically modify the forms of utterances to ease comprehension. In some cases, variation emerges naturally from the manner in which production operates, and then comprehension processes learn to infer the validity of the systematic cue that production provides. For example, there is nothing about positioning a word early in a sentence that inherently marks it as referring to given information; this systematic variation is likely to be a consequence of normal production that comprehension processes become sensitive to. In other cases, an easily comprehended utterance comes from a pressure to create easily produced sentences. For example, speakers will tend to produce a more easily understood common word instead of a less easily understood uncommon word, not because of the difference in comprehension ease *per se*, but because production of the more common word is easier than production of the less common word. Interestingly, a different kind of mechanism seems to underlie the adjustments speakers make to the content of their utterances based on listener knowledge. This is described next.

MODIFICATIONS OF THE CONTENT OF SPOKEN UTTERANCES

Speakers cater the content of their utterances to the comprehension strategies and the knowledge states of their listeners. Interestingly, the role that the actual listener plays in the nature of these listener-sensitive modifications varies; sometimes, the modifications are quite independent of the actual knowledge states of listeners, whereas at other times, speakers specifically take into account what their listeners know. This variability in how listeners' knowledge is accounted for during production seems to reflect a general mechanism that people use when they estimate the knowledge states of others.

For example, many factors affect what kinds of labels speakers use for reference (e.g. referring to a tree outside as 'the tree on the left'). One factor is

that when possible, a speaker will collaborate with their listener to arrive at a suitable label for successful reference, by proposing a label that the listener can approve of, or suggest modifications to. Another important factor is the prior discourse history between a speaker and listener (Brennan and Clark, 1996). If, for example, a speaker has previously referred to 'the creepy tree' with a particular listener, they can use that label again with that listener without renegotiating from scratch. Interestingly, an egocentric bias arises when speakers use their prior discourse history. Whereas it is the case that with a new listener, speakers are relatively less likely to use a label that was used previously with a different listener, the labels that speakers do use with new listeners are nevertheless more similar to the labels that they've previously used (with different listeners), compared to labels that speakers use for the first time. The egocentric bias is revealed by the fact that a speaker's own discourse history, independent of his or her listener's history, influences the nature of the labels that speakers use.

Another example demonstrates how the egocentric bias operates. It has been shown that in story retellings, speakers are more likely to mention atypical instruments of actions (e.g. an icepick for a stabbing event) than typical instruments of actions (a knife for a stabbing event; see Dell and Brown, 1991). This tendency is an elegant reflection of Grice's maxim of quantity: if an idea to be conveyed includes an action that is performed with a highly typical instrument, then the instrument need not be explicitly described, as the listener is likely to infer that particular instrument if necessary. On the other hand, if the action is performed with an atypical instrument, the instrument should be explicitly described, because a listener cannot infer the existence of an atypical instrument. Interestingly, an egocentric bias still arises, in that it has been shown that under some circumstances, whether the speaker knows the listener to already possess knowledge of the instrument does not affect the tendency to especially mention atypical instruments. This is only true, however, with sentences that describe the action itself. So, for example, speakers are as likely to say 'The robber stabbed the victim with the icepick' whether or not their listeners know about the icepick. On the other hand, in subsequent sentences (after the action has been described), speakers are sensitive to listener knowledge. So, a speaker is less likely to say 'Oh, and there was an icepick' when a listener already knows about the icepick, compared to when the listener does not.

These observations hint at a mechanism that turns out to be common in social-cognitive functioning generally (see Nickerson, 1999), as well as in processing related to the production–comprehension interface specifically. The fact that speakers' initial references to an action include mention of atypical instruments equally often regardless of listener knowledge suggests that speakers' initial production decisions are based on egocentric information. That is, utterances are first formulated based on what the speaker knows, relatively independently of what listeners know. This can account for the general egocentric bias that arises in production. Then, the observation that subsequent references to the instrument are sensitive to listener knowledge suggests that after the initial production, speakers are better able to accommodate the differences between what the speaker knows and what their listeners know. Overall, it appears that speakers accommodate the content of their utterances to the knowledge states of listeners with a general two-stage strategy: speaker-centered, egocentric knowledge guides initial utterance formulation, and then listener-specific knowledge is taken into account in subsequent processing.

Why might the production system adopt such a two-stage strategy? For reasons of processing efficiency. In terms of processing resources, it is costly to keep track of the knowledge states of all listeners, with respect to all possible facts that a speaker could express. A reasonable heuristic is that if the speaker knows some fact, his or her listeners are also likely to know that fact. Based on this egocentric strategy, production processes can efficiently formulate an initial utterance that is likely to be good enough for successful communication. However, once the initial utterance is formulated, the specific information expressed in the utterance can more easily be evaluated with respect to listener knowledge. So, the initial formulation can be followed by elaborations and modifications that take into account listener knowledge. Interestingly, if production occurs with relatively little pressure, it is possible that this formulation-then-adjustment strategy could occur prior to articulation. Indeed, this kind of mechanism seems to reflect how common ground information is processed, which is described next.

COMMON GROUND EFFECTS

For a speaker to make definite reference – for example, as occurs with the use of a noun phrase with the definite determiner 'the' – she or he must assume the truth of a remarkable number of facts.

First, the speaker must know about the entity to which she or he wishes to refer. Second, the speaker must know that the listener knows about this entity. And third, the speaker must think that both the speaker him- or herself as well as the listener know about the other's knowledge of this entity. For example, Mary cannot say to Bill 'the guy you voted for is a clown' unless: (a) Bill knows who Bill voted for; (b) Mary knows who Bill voted for; and (c) Mary thinks that both of them know of the other's knowledge of who Bill voted for. Shared knowledge is not sufficient for definite reference, in that it cannot merely be the case that each independently knows of Bill's vote. There must be mutual knowledge, in that both knows of the other's knowledge. If Bill doesn't know that Mary knows of his voting choice, then Mary's statement can fairly be responded to with 'how do you know who I voted for?' It is this mutual knowledge that a set of interlocutors possesses that is termed *common ground*.

THE NATURE OF COMMON GROUND

Without some semblance of the common ground between speaker and listener already in place, a conversation can hardly get off the ground. Many a science fiction story portrays two members of mutually unknown species, struggling futilely to communicate without any shared basis of communication, until that shared basis is discovered, at which point the journey of information exchange begins. The same occurs any time two strangers meet, if less dramatically. In fact, 'current weather conditions' as the reliable kindling of any conversation can be understood in this light: the current weather conditions are encompassing enough that any co-present interlocutors can assume them as a shared basis of conversation (i.e. that it's likely to be common ground), though they are changing and important enough that they are (more or less) worth talking about (thus permitting speakers to respect Grice's maxim of quantity).

Though common ground begins as a necessary starting point for a conversation, it then continues to build across that conversation, and indeed across the continuing interactions between any given interlocutors. Common ground refers not only to what interlocutors can talk about, but how they can talk about it. For example, based on common ground, a speaker will decide whether to speak in English, French, or American Sign Language, and in those languages, what register to use. Research has identified two different bases that jointly contribute to these many different aspects of the common ground that is shared between interlocutors (Clark, 1996). Communal common ground refers to the abilities and knowledge that interlocutors can assume to be possessed by virtue of the community membership of the participating interlocutors. Thus, for example, if a speaker is aware that his or her interlocutor is a French Canadian, an accountant, a bridge player, or a parent, she or he will assume mutual knowledge of a set of facts that members of each of those communities will typically know. Personal common ground refers to the set of facts that interlocutors come to know as a consequence of their joint experiences, including things that happen within a conversation (the entities that speakers refer to) as well as those outside the conversation (events that interlocutors can be assumed to have noticed). All facts learned in past conversations or mutual experiences among a set of interlocutors become parts of those interlocutors' common ground.

PROCESSING CONSIDERATIONS

Speakers clearly recognize information that is in common ground when they speak. This is evident not only from the fact that speakers' definite references overwhelmingly succeed in conversation, but also by more subtle audience design effects (e.g. Krauss and Fussell, 1991), where speakers seem to cater their utterances to account for what their listeners are likely to know already (i.e. for what is in common ground). For example, speakers will refer to local landmarks or provide navigation directions differently if they think their listeners are local residents, compared to if they think the listeners are not local. Speakers will also produce utterances differently if they think their speech is for themselves (i.e. for the purpose of future reference) compared to if the speech is for others, or if speech is intended for a personal friend compared to if it is intended for a stranger. Furthermore, research has shown that communication is more successful when speakers correctly know who their intended audience is.

How speakers recognize common ground is not a trivial issue, however. Because language users do not have direct access to the cognitive states of others, each must infer whether a particular fact is mutually known. There are many specific strategies that interlocutors can use to infer whether a fact is in common ground. For example, if interlocutors notice that each has made eye contact with a particular object in the environment, they may assume that object to be in common ground. Or, if one interlocutor verbally refers to an entity, all

interlocutors who could reasonably be assumed to have heard the verbal reference can have that entity ascribed to their common ground. Interestingly, the egocentric bias described above arises too when speakers attempt to account for common ground. With respect to communal common ground, it has been found that while speakers can quite accurately estimate what other members of specified communities do and do not know, they still exhibit a systematic bias to overestimate the likelihood that others know things that they themselves know. This effect, a version of the 'false consensus effect' from social psychology, suggests that with respect to common ground, an interlocutor who idiosyncratically knows a fact is relatively likely to ascribe knowledge of that fact to others, and therefore assign that fact to common ground.

Research investigating the use of personal common ground has found evidence not only of an egocentric bias, but also for the two-stage model described above. If speakers formulate utterances freely, without any time pressure, they can easily provide enough information to refer successfully to an intended concept in its context (see Deutsch and Pechmann, 1982), and can distinguish information in common ground from privileged information (information that only the speaker knows about). However, if speakers formulate utterances under time pressure, the distinction between common ground and privileged information erodes, as speakers are more likely to make (infelicitous) references to privileged information in utterances that are directed to their listeners. Thus, initial formulations – ones that are used when time pressure does not permit subsequent revision – are based on egocentric information, until subsequent modifications of the initial formulation can correct for the different knowledge states that speakers and listeners might have (see Horton and Keysar, 1996).

Overall, then, for communication to succeed, speakers must take into account the linguistic capabilities and the knowledge states of their listeners. Grice's cooperative principle provides an ideal to which speakers and listeners can appeal with each contribution to a conversation. Given this ideal, the extent to which speakers follow the cooperative principle is impressive. Although speakers show some general deviations from producing language in an entirely cooperative manner, these can largely be understood as the result of speakers' use of fallible heuristics – heuristics that are necessary given the pressures involved in timely production, as well as the intractability of the demands involved in estimating mutual knowledge. Despite these difficulties, speakers modify the form and the content of their utterances so that their listeners can more easily understand those utterances. The listener-sensitive accommodations that speakers make during language production are a revealing compromise between what speakers know, and what they guess others know.

References

Brennan SE and Clark HH (1996) Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition* **22**: 1482–1493.

Clark HH (1996) *Using Language*. Cambridge, UK: Cambridge University Press.

Cutler A (1987) Speaking for listening. In: Allport A, MacKay DG, Prinz W and Scheerer E (eds) *Language Perception and Production: Relationships between Listening, Speaking, Reading, and Writing*, pp. 23–40. London: Academic Press.

Dell GS and Brown PM (1991) Mechanisms for listener-adaptation in language production: limiting the role of the 'model of the listener'. In: Napoli DJ and Kegl JA (eds) *Bridges between Psychology and Linguistics: A Swarthmore Festschrift for Lila Gleitman*, pp. 105–129. Hillsdale, NJ: Lawrence Erlbaum Associates.

Deutsch W and Pechmann T (1982) Social interaction and the development of definite descriptions. *Cognition* **11**: 159–184.

Fox Tree JE and Clark HH (1997) Pronouncing 'the' as 'thee' to signal problems in speaking. *Cognition* **62**: 151–167.

Grice HP (1975) Logic and conversation. In: Cole P and Morgan JL (eds) *Syntax and Semantics*, vol. 3: *Speech Acts*, pp. 41–58. New York, NY: Academic Press.

Horton WS and Keysar B (1996) When do speakers take into account common ground? *Cognition* **59**: 91–117.

Krauss RM and Fussell SR (1991) Perspective-taking in communication: representations of others' knowledge in reference. *Social Cognition* **9**: 2–24.

Nickerson RS (1999) How we know – and sometimes misjudge – what others know: imputing one's own knowledge to others. *Psychological Bulletin* **125**: 737–759.

Further Reading

Bock K (1995) Sentence production: from mind to mouth. In: Miller JL and Eimas PD (eds) *Handbook of Perception and Cognition*, vol. II: *Speech, Language, and Communication*, pp. 181–216. San Diego, CA: Academic Press.

Brown PM and Dell GS (1987) Adapting production to comprehension: the explicit mention of instruments. *Cognitive Psychology* **19**: 441–472.

Clark HH and Clark EV (1977) *Psychology and Language: An Introduction to Psycholinguistics*. New York, NY: Harcourt Brace Jovanovich.

Clark HH and Marshall CR (1981) Definite reference and mutual knowledge. In: Joshki AK, Webber BL and Sag IA (eds) *Elements of Discourse Understanding*,

pp. 10–63. Cambridge, UK: Cambridge University Press.

Ferreira VS and Dell GS (2000) Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology* **40**(4): 296–340.

Levelt WJM (1983) Monitoring and self-repair in speech. *Cognition* **14**: 41–104.

Levelt WJM (1989) *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

MacKay DG (1987) *The Organization of Perception and Action: A Theory for Language and other Cognitive Skills*. New York, NY: Springer-Verlag.

Newport EL, Gleitman H and Gleitman LR (1977) 'Mother, I'd rather do it myself': some effects and non-effects of maternal speech style. In: Snow CE and Ferguson CA (eds) *Talking to Children: Language Input and Language Acquisition*. Cambridge, UK: Cambridge University Press.