6 Parity and Disparity in Conversational Interaction

JENNIFER S. PARDO

Montclair State University

Introduction

Conversation in social interaction is a primary site for language use. However, Chomsky's emphasis on competence over performance set the mainstream agenda for early research in psycholinguistics, which focused on relatively encapsulated, mechanistic approaches to language. By separating language from culture, researchers made progress on specific questions about relationships between language and cognition. Of course, this revolutionary approach fostered a dialectical turn to research on social aspects of language use (Searle, 1976). In that vein, early research by luminaries such as Robert Krauss, Herb Clark, and Howard Giles demonstrated a profound influence of the target audience on language use in social interaction. While communicating linguistic messages, utterances also serve as a medium for nonlinguistic signals about a talker's identity and orientation to an addressee and their situation. This interplay evokes both parity and disparity across multiple aspects of conversational interaction. In some ways, talkers become more similar, but at the same time, they maintain and communicate their individual identities. Thus parity, while useful for mutual understanding of a linguistic message, is not necessarily carried through to every facet of spoken communication.

Conversation can be messy and unpredictable, making it an impractical target for rigorous scientific investigation. Controlled studies have introduced constraints on free-form conversation by using interviews, referential communication tasks, or completely constrained communication frames (such as semi-scripted sentence prompts). The upshot is that the form and function of a message is influenced in dramatic ways by attributes of a target addressee and social settings of language use. Accordingly, "the meaning of a message is more usefully thought of as something that is negotiated between a sender and receiver" than as something completely determined by a sender's aim (Krauss 1987, p. 86). Acknowledging the

The Handbook of Psycholinguistics, First Edition. Edited by Eva M. Fernández and Helen Smith Cairns.

importance of an addressee in spoken communication departs from traditional psycholinguistic conceptualizations of communication as message transmission from sender to receiver, but a complete appreciation for language use in its most natural setting demands that the field take this leap.

The following sections survey research on conversational interaction starting with early work from Robert Krauss' *Perspective-Taking* and Herb Clark's *Coordination* approaches to language use. Both approaches emphasize language use as a coordinated activity among interlocutors who attempt to build shared meaning. In a roughly parallel vein, Howard Giles and colleagues developed *Communication Accommodation Theory* (CAT), which explores social constraints on language variation during conversational interaction. This approach was driven by findings that talkers code-switch between dialectal variants in a way that often increases their similarity to an interlocutor's dialect. That is, a talker will adjust their pronunciation and vocabulary to approach or even match that of their target audience. While CAT enumerated many social constraints on speech accommodation, recent research on alignment/convergence in dialogue reveals the importance of these phenomena for constraining basic cognitive mechanisms of language use in social interaction. The final section offers a framework for tying these approaches together under a general entrainment perspective on coordinated behavior.

Perspective-taking and coordination in language use

Some of the earliest studies demonstrating perspective-taking in referential communication were conducted by Robert Krauss and colleagues (and beautifully summarized in Krauss, 1987). These studies indicate that expectations about a target addressee influence the form of referential descriptions in ways that have implications for their communicative function. The basic paradigm involves a matching game with sets of cards depicting novel ambiguous figures that each evoke a complex descriptive utterance. This paradigm has been used both in non-interactive settings with hypothetical target addressees, as well as in task-oriented conversational studies.

In a series of non-interactive studies, talkers were asked to provide descriptions of ambiguous figures that could be used to identify them at a later time (Fussell & Krauss, 1989; Krauss, 1987; Krauss, Vivekananathan, & Weinheimer, 1968). In order to examine the impact of perspective-taking on message formulation, talkers provided separate descriptions of these figures for themselves, for a friend, and for a stranger. With regard to content, messages intended for strangers were longer, and comprised less diverse vocabulary and more literal descriptions than messages for the self. With regard to accuracy in matching the items with their descriptions, messages intended for the self yielded better performance than those for strangers, despite being shorter and less generic. Finally, when a target addressee was a specific friend, matching accuracy fell between that of messages for the self and for a stranger and was influenced by degree of friendship (Krauss, 1987). As a talker's knowledge of a target addressee increased (from stranger to friend to self in this

case), descriptions became shorter and less literal, as well as more obscure to those who didn't share common ground with a talker.

When these figures are used in task-oriented conversational studies, talkers actively engage with target addressees in a matching task. In these tasks, each member of a pair of interlocutors receives a set of figures, but their figures are in different orders. One member is designated the Director in the task and the other member is the Matcher. The goal is for interlocutors to converse so that a Matcher can place their figures in the order dictated by a Director's set. Because the figures are ambiguous, the main activity of the task revolves around development of referential terms for the figures. As interlocutors played multiple rounds of the matching game with the same set of figures in different orders, descriptions of figures became shorter and more figurative, and matching accuracy increased, similar to the pattern found when comparing descriptions for strangers with those for friends and the self in non-interactive studies (Krauss, Bricker, Garlock, & MacMahon, 1977; Clark & Wilkes-Gibbs, 1986; Krauss & Bricker, 1967; Krauss & Weinheimer, 1964). Furthermore, degree of reduction in message length was influenced by whether or not a Matcher could provide verbal feedback to a Director at all (Krauss & Weinheimer, 1966), and whether feedback was delayed (Krauss & Bricker, 1967), indicating that Matchers actively participate in a process of referential description tailoring.

This tailoring by Matchers was demonstrated most dramatically in an interactive conversational study by Schober and Clark (1989; see also Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs & Clark, 1992). In their study, they contrasted an autonomous view of communication as a series of individual acts of language production and comprehension with a view that communication involves the accumulation of common ground between interlocutors through an active process of grounding (i.e., coordination), in which addressees tailor messages to be specifically designed for them. The study involved 10 pairs of talkers playing a matching game with ambiguous figures that were recycled over the course of six trials. On each trial, a Director and Matcher both had 16 figures, and the Director instructed the Matcher in placing 12 of them in a pre-determined order. This design permitted measures of number of words used to identify each figure, accuracy in figure identification, and timing of figure identification. In a crucial twist, the recorded conversations were presented to 40 additional Matchers (dubbed Overhearers), who completed the matching task in a non-interactive manner. Half of the Overhearers had access to entire conversations (Early Overhearers), while the other half could only access recordings from the third trial on (Late Overhearers).

Taking a viewpoint that language occurs in autonomous production and comprehension steps, there should be no difference in matching accuracy between (active) Matchers and Early Overhearers who had access to an entire conversation. Because both Matchers and Early Overhearers were strangers to Directors at the outset, the first referential descriptions were presumably designed for any generic stranger, and Early Overhearers had access to the same content as Matchers, including their feedback, questions, and so on. If common ground in communication merely accumulates via distinct acts of language production and comprehension, then shared content should yield equivalent performance between Early

Overhearers and Matchers. However, if, as Schober and Clark hypothesized, common ground is generated through an active process of grounding between interlocutors, with Matchers evoking messages that are specifically designed for them, then Early Overhearers should not perform as well as active Matchers.

During the task, the number of words per figure uttered by Directors started at an average of 73 in trial 1 and decreased steadily to asymptote at around 13 by trial 4. Likewise, the number of turns and amount of time spent per figure decreased from trial 1 to trial 6. This pattern contrasted with a slight increase in figure identification accuracy for both Matchers and Overhearers over the course of the trials. However, Matchers outperformed Early Overhearers in figure identification on all trials, with an average of 99% versus 88% correct identification, and Late Overhearers (who started with trial 3) fared even worse at 68% correct (chance performance could be as high as 20% for the last figure in a trial, presuming it was not already misidentified). An additional manipulation enabled some Overhearers to pause conversation recordings while completing the task, but their performance was no different from those who could not pause the recordings. With respect to timing of figure identification, Overhearers were much more variable relative to Matchers. Finally, these patterns were replicated in a second experiment in which Overhearers were present during the live exchanges, but unable to collaborate directly with the interacting talkers. According to Clark and Wilkes-Gibbs (Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs & Clark, 1992), both members of a conversing pair negotiate shared meaning through a process involving presentation, acceptance/rejection, and mutual recognition phases, which serve to minimize collaborative effort in grounding reference.

These findings challenge a traditional conceptualization of communication as message transmission from a sender to a receiver via an informational medium. According to the encoding-decoding paradigm, the sender is the cause of a message that a receiver simply decodes (see Krauss & Chiu, 1998; Krauss & Fussell, 1996). By adopting this paradigm, one can justify separating communication into distinct acts of language production and language perception that can be analyzed and studied separately. The organization of the field of psycholinguistics into language production and comprehension implicitly validates this perspective. However, research on language use in conversation belies this division by demonstrating that the form and efficacy of a message depend on expected and actual knowledge states of an intended audience. "[T]he traditional separation of the roles of participants in a verbal interaction into sender and receiver, speaker and addressee, is based on an illusion—namely that the message somehow 'belongs to' the speaker, that he or she is exclusively responsible for having generated it, and that the addressee is more-or-less a passive spectator to the event…. [T]he message, in the concrete and particular form it takes, is as much attributable to the existence of the addressee as it is to the existence of the speaker" (Krauss, 1987, p. 96).

These foundational studies demonstrated the importance of an addressee in message formulation, eliciting a turn to an ongoing round of research examining implications of perspective-taking and coordination for cognitive mechanisms of language use. For example, Schober (1993) demonstrated that talkers in a task

involving differing spatial perspectives tended to use more egocentric frames of reference when interacting with a partner than when alone, presumably due to the potential for a live partner to provide feedback and/or to solicit clarification as needed. In more recent studies, the scope of referential communication has expanded beyond development of terms for ambiguous figures to repeated reference with more familiar items. In these studies, the set of objects used in a matching game can elicit basic level terms (such as *candle* or *shoe*) when objects span different categories, or more specific terms (such as *small candle* or *loafer*) when objects appear with category competitors (e.g., Barr & Keysar, 2002; Brennan & Clark, 1996; Keysar, Barr, Balin, & Brauner, 2000).

An overarching issue in such research has been when and how a target addressee's perspective is incorporated during utterance planning, and whether addressees likewise interpret utterances against common ground. Needless to say, there is ample evidence supporting accounts that favor an early single-stage model in which an addressee's perspective is taken into account immediately and referential terms are negotiated (e.g., Brennan, Galati, & Kuhlen, 2010), as well as those that propose a dual-process model in which primary egocentrism is supplemented by secondary perspective adjustment (e.g., Barr & Keysar, 2006). As reviewed in the previously cited papers (and in Schober & Brennan, 2003), sometimes interlocutors fail to take a partner's perspective into account (Bard *et al*., 2000; Horton & Keysar, 1996), and sometimes what appears to be an addressee influence might be a coincidence of shared perspective (Brown & Dell, 1987; Kesyar *et al*., 2000).

Regardless of when and how an addressee influences message formulation, this process can evoke increased similarity between interlocutors in a number of dimensions, sometimes referred to as communication accommodation (e.g., Giles, Coupland, & Coupland, 1991; Shepard, Giles, & Le Poire, 2001), or as alignment between interlocutors (e.g., Pickering & Garrod, 2004, 2007, 2013). In general, Giles' communication accommodation perspective focuses on social factors driving both convergence and divergence of talkers, while Pickering and Garrod's alignment perspective focuses on internal cognitive mechanisms that promote parity among interlocutors across all facets of language.

Communication accommodation

Studies of interacting talkers have found fairly consistent patterns of linguistic change over the course of conversational interaction, and such changes are variously termed *coordination* (Branigan, Pickering & Cleland, 2000; Brennan & Clark, 1996; Clark, 1996), *entrainment* (Arantes & Barbosa, 2010; Brennan & Clark, 1996; Levitan & Hirschberg, 2011; McGarva & Warner, 2003), *alignment* (Branigan, Pickering, McLean, & Cleland, 2007; Pickering & Garrod, 2004, 2007, 2013), or *accommodation* (Giles *et al*., 1991; Shepard *et al*., 2001). Most of these proposals examine increasing similarity (convergence) of diverse aspects of interlocutor's speech, from schematic (Garrod & Doherty, 1994), to syntactic (Branigan *et al*., 2000; Branigan *et al*., 2007), to lexical/semantic levels (Brennan & Clark, 1996; Fusaroli,

Bahrami, Olsen, Roepstorff, Rees, Frith, & Tylén, 2012; Krauss & Weinheimer, 1964; Nenkova, Gravano, & Hirschberg, 2008; Neiderhoffer & Pennebaker, 2002; Wilkes-Gibbs & Clark, 1992). Research on convergence has also included measures of acoustic attributes such as perceived accentedness, sub-vocal spectral covariation, and voice amplitude (e.g., Giles, 1973; Gregory & Webster 1999; Heldner, Edlund, & Hirschberg, 2010; Levitan & Hirschberg, 2011; Natale, 1975). Convergence in such parameters appears to be influenced by social factors that are local to communication exchanges, such as interlocutors' relative dominance or perceived prestige (Gregory, Dagan, & Webster, 1997; Gregory & Webster, 1996).

Howard Giles's *Communication Accommodation Theory* (CAT) also acknowledges the opposite pattern, accent divergence, under some circumstances (Giles *et al*., 1991; Shepard *et al*., 2001). Although it is tempting to attribute convergence to an automatic imitative function that increases intelligibility for the parties involved (e.g., Pickering & Garrod, 2004), divergence often does not preclude intelligibility, but serves a communicative purpose for a diverging party (Bilous & Krauss, 1988; Bourhis & Giles, 1977; Labov, 1974). It is arguable that convergence likewise serves a communicative purpose beyond intelligibility. One reason proposed for accommodation is the similarity attraction hypothesis, which claims that individuals try to be more similar to those to whom they are attracted (Byrne, 1971). Accordingly, convergence arises from a need to gain approval from an interacting partner (Street, 1982) and/or from a desire to ensure smooth conversational interaction (Gallois, Giles, Jones, Cargile, & Ota, 1995). Divergence is often interpreted as a means to accentuate individual/cultural differences or to display disdain (Bourhis & Giles, 1977; Shepard *et al*., 2001).

According to CAT, talkers also converge or diverge along different speech dimensions as a function of their relative status or dominance in an interaction (Giles *et al*., 1991; Jones, Gallois, Callan, & Barker, 1999; Shepard *et al*., 2001), which is compatible with the similarity attraction hypothesis. Typically, a talker in a less dominant role will converge toward a more dominant partner's speaking style (Giles, 1973). In contrast, a talker's speech might diverge from a conversational partner's to accentuate distinctiveness, regardless of dominance (Bourhis & Giles, 1977). Finally, talkers have been found to converge on some parameters at the same time that they diverge on others (Bilous & Krauss, 1988).

In an early empirical study by Giles (1973), a group of Bristol men were interviewed by two different interviewers—one spoke with a prestigious Received Pronunciation (RP) accent, and the other was a Bristol-accented interviewer. Excerpts from the interviewees' recordings were played to separate Bristol listeners, who rated how Bristol-accented the speech sounded. When talkers interacted with the RP-accented talker, their speech was rated as less Bristol-accented than when interacting with the Bristol-accented talker. Giles interpreted this pattern as accent convergence toward the RP interviewer (i.e., upward convergence to a higher-status individual).

A second study with Welsh men found accent divergence (Bourhis & Giles, 1977). In this case, Welsh-accented talkers answered pre-recorded interview questions spoken by a single RP-accented talker. During a break in the interview, the

talkers overheard the RP-accented talker make some disparaging remarks about the Welsh language (ending with, "the future of Welsh appears pretty dismal," p. 125). Then, the talkers answered a second set of questions recorded from the same RP-accented talker, and excerpts were presented to separate listeners, this time comparing ratings from the pre-insult phase to those for speech produced after the insult. In this case, talkers increased their Welsh-accentedness in the post-insult phase, displaying their Welsh status by diverging from the RP-accented talker. In one case, a Welsh speaker refused to answer questions in the post-insult phase, and responded to every question by conjugating Welsh verbs.

These early studies demonstrated both convergence and divergence in code-switching behavior that formed the basis for a fruitful line of research on social and situational modulators of communication accommodation (summarized in Giles *et al*., 1991; Shepard *et al*., 2001). While most studies have focused on a single speech attribute (such as speech rate or accentedness), a crucial study by Bilous and Krauss (1988) on the influence of talker sex on accommodation demonstrated that the landscape of accommodation is extremely complex. Their study assessed the so-called male dominance hypothesis with regard to convergence by comparing baseline measures collected when talkers interacted in same-sex pairs to those when they interacted in mixed-sex pairs (in a total of 60 talkers, half male). Accordingly, when interacting in mixed-sex pairs, females should converge to male speaking patterns, while males should not change their patterns when interacting with females.

Comparing across multiple measures in same- and mixed-sex pairings, Bilous and Krauss (1988) found that convergence in mixed-sex pairs was neither consistent across measures, nor were patterns of convergence explained by the male dominance hypothesis. Men and women both converged in average utterance length and frequency of short and long pauses. Women converged to men in total number of words and in frequency of interruptions, and diverged from men in frequency of back-channels and frequency of interruptions. Men converged to women in frequency of back-channels and frequency of laughter and did not diverge in any measures.

The results reported by Bilous and Krauss (1988) provided an early indication that convergence is not an all-or-none phenomenon. Talkers may converge on some attributes at the same time that they diverge on others. In this case, they measured structural attributes of conversational interaction that related to conversational dominance (holding the floor, managing turn-taking, etc.). These findings are important because they point to a major difficulty in measuring convergence. That is, no single measure can provide a comprehensive assessment of convergence in social interaction. This point will be elaborated further in a discussion of measures of phonetic convergence.

Interactive alignment

It is clear that aspects of a social/cultural setting and relationship between interlocutors will influence the form and direction of communication accommodation. However, much of the research within the accommodation framework is mute

regarding internal cognitive mechanisms that support convergence and divergence during speech production, which have only recently been taken up in the field of psycholinguistics. All such accounts of the phenomenon rest on an assumption of parity of representation between talkers and listeners. In order for a listener to converge to a talker, they must create a sufficiently detailed representation of the talker's speech. This assumption plays a central role in Pickering and Garrod's interactive alignment framework for dialogue (Pickering & Garrod, 2004, 2007, 2013).

In their mechanistic approach, Pickering and Garrod (2004) proposed a model of language use in dialogue based on a simple idea. That is, automatic priming of shared representations leads to alignment at all levels of language—semantic, syntactic, and phonological. Moreover, alignment at one level promotes alignment at other levels. On those occasions when the default automatic priming mechanism fails to yield schematic alignment (e.g., during a misunderstanding), a second more deliberate mechanism brings interlocutors into alignment. The proposed model supports inclusion of an automatic priming mechanism by citing evidence for between-talker alignment at semantic, syntactic, and phonological levels.

In their most recent paper, Pickering and Garrod (2013) draw out a critical component of the model—that language production and language comprehension processes are tightly interwoven within talkers due to processes entailed in self-monitoring of speech production. In particular, they extend concepts from theories of more general action production and perception to language. To do so, they rely heavily on the notion of efference copy, which developed in classic approaches to visual perception (in particular for accommodating the impact of eye movements on motion perception). An efference copy is a secondary signal that is generated along with a motor command (an efference). A separate monitoring system uses an efference copy signal to generate an expected sensory outcome (a forward model) that can be compared to an organism's actual sensory signal resulting from an action. Accordingly, self-monitoring during language production involves a parallel forward modeling component that uses efference copies of speech motor commands to generate simulated perceptual consequences of production. These same forward modeling processes used in speech production also drive active simulation during speech perception, which then leads to covert and sometimes overt imitation. Therefore, when a listener hears an utterance, comprehension relies on the same processes as production, leading to convergent production in a very straightforward manner. A more elaborate account of this model linking language production and perception appears in this volume (Gambi & Pickering, this volume; see also Gambi & Pickering, 2013).

Evidence in support of semantic/schematic alignment is abundant, and is to be expected, because communication could hardly be successful if interlocutors failed to use similar terms. Much of the previously discussed research on development of referential terms supports a proposal of semantic alignment. In situations investigated so far, interlocutors aligned on terms to refer to relevant properties of a task. In other less constrained situations, however, it is acknowledged that some

degree of disparity in representations can be tolerated. When disparity is too great, interlocutors then must negotiate to restore mutual understanding (Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs & Clark, 1992).

A compelling study by Garrod and Doherty (1994) extended the concept of schematic alignment within an individual dyad to alignment across members of a closed community of talkers who each interacted with every other member of the community. Participants in their study played a computerized maze game in which each member of a pair moved an icon from a starting cell to an ending cell in their maze. The structure of the game required conversational collaboration, in that each interlocutor had to guide their partner to particular locations in a maze in order to complete their own movements. During task completion, interlocutors used multiple descriptions schemes for mazes that became aligned over the course of interaction. For example, one talker might use a scheme that involved referring to maze cells according to a matrix notation, while another might use a scheme involving path-like descriptions (there were four basic description schemes). Eventually, both talkers would use the same scheme on adjacent turns.

Each individual played the maze game nine times, with a limit of 10 minutes/game. A total of 20 talkers participated, either as individual pairs or as a closed community (N = 10 each). The individual pairs played the game nine times with the same partner. The closed community group played the game nine times as well, but each game was played with a different partner so that by the end, each person had played the game only once with every other person in the group. Alignment was measured as between-talker consistency in using the same description scheme from one turn to the next.

Individual pairs rapidly converged on consistent schemes in early games, and persisted in the same high level of consistency throughout the interaction. The community group started out with lower levels of consistency than individual pairs, but by the middle and late games (games 4-9), their levels of consistency were actually higher than that of individual pairs, despite the fact that they changed partners on every game. A second experiment ruled out the possibility that changing partners alone caused the increased consistency—what mattered was that the talkers formed a closed community, with each member of the community carrying the most frequent scheme from one interaction to the next. Thus, schematic alignment occurs within individual pair-wise interactions, but can also become more consistent across a community of mutually interacting talkers.

Semantic alignment is central to notions of communicative efficacy. Interlocutors do not reach mutual understanding without some degree of parity or shared meaning, which often results in shared lexical forms. However, syntax is demonstrably separate from semantics, therefore, parity in communication does not entail parity in syntactic form. With respect to a situation model, it is irrelevant whether one chooses to say, *The pirate gave the banana to the clown*, or, *The pirate gave the clown the banana*. Both structures yield the same understanding of the expressed situation, despite their syntactic differences. Although communicative efficacy does not necessarily drive syntactic parity, early corpus studies found that the likelihood of producing a particular syntactic form was greater if that same

variant had been used previously in a discourse (Estival, 1985; Levelt & Kelter, 1982; Weiner & Labov, 1983). These findings, which came to be known as syntactic repetition/priming, led to a host of laboratory studies of the phenomenon, beginning with Bock (1986) and well summarized by Hartsuiker, Bernolet, Schoonbaert, and Vanderelst (2008) and Gries (2005). Overall, syntactic repetition has been found in situations without repetition of lexical items, but the findings are stronger with concurrent repetition of lexical forms and in dialogue tasks (Hartsuiker *et al*., 2008), and are generally similar to those reported in uncontrolled corpus studies (Gries, 2005).

Two studies by Branigan and colleagues examined syntactic repetition in controlled dialogue, providing rigorous support for the proposal that interlocutors might align on syntactic forms (Branigan *et al*., 2000; Branigan *et al*., 2007). In these studies, a talker interacted with a confederate in an alternating picture description and matching game—one talker described a picture to be matched by their partner, and then they switched roles on the next picture. The pictures comprised line drawings of an agent holding an object next to a recipient of the object, and the describers' cards were labeled with a verb that was to be used in the description (there were six different verbs, and each verb appeared on two cards).

These materials elicit sentences with a syntactic frame that could alternate between a Prepositional Object (PO: the agent verbing the x *to* the y) or a Double Object (DO: the agent verbing the y the x). Confederates were provided with scripted sentences that controlled whether they used the PO or DO form on each trial. Of interest was whether the participants would use the same form or the alternate form on their subsequent turn describing a new picture. Note that these conversations were heavily constrained such that participants could only repeat the same descriptions if clarification was needed.

Branigan *et al*. (2000) found evidence that talkers (N = 24) were more likely to use the same syntactic frame as on a previous trial, and that the effect was stronger when the verb was also the same. Participants were 55% more likely to use the same syntactic form as the confederates when the verbs were the same, and 26% more likely when the verbs differed. Crucially, the items depicted were not the same, so the effect was not due to overlap in other forms of sentence content. For example, if the confederate described *the cowboy handing the banana to the burglar*, the participant would then describe *the pirate handing the cake to the sailor* (as opposed to *the pirate handing the sailor the cake*).

Branigan *et al*. (2007) extended these findings to investigate whether the relationship between the confederate and the participant mattered. In this case, there were three talkers, and the confederate could have been addressing the participant directly or another talker prior to the participant's description. When a participant had been the addressee of the prior description, they were 32% more likely to use the same form in their subsequent description. When they had merely been a side-participant to the confederate's utterance (they were told to check the other talker's descriptions), they were only 12% more likely to use the same form. These results lend some support to the notion that interlocutors often converge on the same syntactic forms during conversational interaction, but note that talkers often

used the other form, indicating that activation of particular syntactic procedures does not fully determine the syntactic form of an utterance.

At the phonological level, a growing body of research on phonetic convergence both supports and challenges a mechanistic approach based on automatic priming. On one hand, studies have found that interacting talkers become more similar in phonetic repertoire, whether talkers come from the same or different dialect regions. On the other, observed patterns of phonetic convergence are not readily accommodated by a framework that relies so heavily on an automatic priming mechanism. To date, most of the psycholinguistic research on phonetic convergence has used non-interactive speech shadowing tasks (e.g., Goldinger, 1998), passive exposure tasks (e.g., Nielsen, 2011), or examined the impact of accumulated exposure (e.g., Pardo, Gibbons, Suppes, & Krauss, 2012; Sancier & Fowler, 1997; see review in Pardo, Jordan, Mallari, Scanlon, & Lewandowski, 2013). In studies of phonetic convergence during conversational interaction, talkers have been found to converge, but effects are subtle and uneven across multiple measures. Most importantly, as in the Branigan *et al*. (2007) study of syntactic alignment, phonetic convergence is dramatically influenced by a talker's role in a conversational setting.

Phonetic convergence during conversational interaction

Inspired by a finding that talkers "imitated" model speech prompts in a non-interactive speech shadowing task (Goldinger, 1998), Pardo (2000, 2006) adapted the paradigm to examine phonetic convergence during conversational interaction. In addition to establishing phonetic convergence during conversational interaction, Pardo (2006) found that a talker's role in a conversation and the sex of the pair of talkers both influenced degree of phonetic convergence. Subsequent studies explored the impact of role by manipulating a talker's intention to imitate (Pardo, Cajori Jay, & Krauss, 2010) and role stability (Pardo, Cajori Jay, Hoshino, Hasbun, Sowemimo-Coker, & Krauss, 2013). In all studies, it was necessary to introduce a constrained conversational task that would guarantee between-talker repetitions of the same lexical items.

To obtain samples of naturalistic conversational speech with appropriate repetitions, the paradigm employed a modified version of the Map Task, a cooperative conversational task that was developed by the Human Communication Research Center at the University of Edinburgh, Scotland (Anderson *et al*., 1991). The Map Task comprises paired maps with labeled iconic landmarks (e.g., *walled city*, *wheat field*, *green bay*). One map in each pair, designated the Giver's map, has a path drawn from a starting point, around various labeled landmarks, to a finishing point. The corresponding map, designated the Receiver's map, has only a starting point and various labeled landmarks. The goal of the task is for a pair of talkers to communicate effectively enough that the Receiver can duplicate the path that is drawn on the Giver's map without seeing each others' maps. The Map Task is particularly useful for studying phonetic convergence during conversational interaction because talkers naturally repeat the landmark labels, and changes in an individual's phonetic

repertoire can be assessed within the same lexical items by collecting recordings of the landmark label phrases before and after conversational interaction.

Measures of phonetic convergence included both acoustic analyses of speech samples and an AXB perceptual similarity test that compares a talker's pre-task and task utterances (A/B) to their partner's utterances (X) (adapted from Goldinger, 1998). This use of an AXB perceptual similarity task ensures that the apparent similarity between talkers was not coincidental—the measure reflects change in a talker's phonetic repertoire that makes them sound more similar to a model than they were prior to exposure. Measuring phonetic convergence in a perceptual task is preferable to phonetic transcriptions due to difficulties entailed in obtaining reliable transcriptions, and to ensure that the measure reflects changes that are available to ordinary listeners. Moreover, perceptual assessment provides a holistic appraisal of similarity that integrates over all acoustic-phonetic dimensions, avoiding potential pitfalls involved in committing to a single acoustic measure (Pardo, 2013; Pardo, Jordan *et al*., 2013; Pardo & Remez, 2006). It is likely that different pairs of talkers might converge on distinct acoustic-phonetic attributes, and perceptual measures of similarity reflect patterns that would be missed when measuring acoustic attributes alone. Finally, identification of converging acoustic attributes alone leaves open the question of whether such attributes are perceptually salient and available for use during conversational interaction.

Across multiple studies of conversational interaction, talkers were found to converge in phonetic form. That is, a talker's utterances of landmark label phrases, such as *green bay* or *diamond mine*, sounded more similar to their partner's utterances during or after conversational interaction than before the talkers met. As in studies of syntactic convergence, measures of phonetic convergence indicate that the change is subtle and variable. For example, listeners selected the conversational task items as more similar to the partner's items on 65% of trials in one study (Pardo, 2006), but on only 53% of trials in another study (Pardo *et al*., 2010). Both findings were significantly greater than chance responding (50%), but the overall range of values reported across multiple studies indicate that the phenomenon is often subtle and highly variable.

Overall, there is a great deal of variability in phonetic convergence across individual pairings that is not readily explained in these studies and merits further research. Across studies of phonetic convergence, measures for individual talkers ranged from 33% to 83% detected convergence in AXB tasks. However, in contrast with earlier findings from the communication accommodation literature, talkers in these studies did not converge in speaking rate, presumably due to an influence of role discrepancy in the Map Task (Pardo *et al*., 2010; Pardo *et al*., 2013). That is, Givers tended to speak faster than Receivers, maintaining a significant difference in speaking rate that Receivers neither matched nor tracked in their own speaking rates. Furthermore, item and vowel analyses have found inconsistent patterns of vowel convergence across pairs—individual pairs converged on unique acoustic-phonetic attributes that were apparent to listeners who made a global perceptual appraisal (see Pardo, Jordan *et al*., 2013).

Taken together, these findings illustrate the complexity of phonetic convergence in conversational interaction. A glance at the literature on phonetic convergence in more constrained non-interactive settings is no less complex (see reviews in Pardo,

Jordan *et al*., 2013; Pardo, Urmanche, Wilman, & Wiener, 2017). Although it is clear that some form of representational similarity must underlie a talker's ability to converge in phonetic form to an interlocutor, it is not clear how this similarity arises and how it interfaces with a talker's appraisal of their social setting to determine their degree of convergence. Automatic priming mechanisms are not sufficient to explain simultaneous findings of global phonetic convergence and inconsistent patterns in individual acoustic attributes such as articulation rates and vowel formants. Such findings can be accommodated within the framework of entrainment, which explicitly incorporates both patterns of convergence and divergence among interacting dynamical systems.

Entrainment

The principles of entrainment, the *maintenance tendency, superimposition*, and the *magnet effect*, were initially identified in von Holst's (1937/1973 in Gallistel, 1980) early research on endogenous rhythmicity in behavioral organization. Examining fish fin oscillations, von Holst discovered endogenous neural oscillators that likely serve as basic building blocks of complex behaviors through superimposition of coordinated patterns (Gallistel, 1981; Turvey, 1990). The *maintenance tendency* describes the observation that each oscillator prefers to operate according to its own intrinsic dynamics (frequency and amplitude). From a basic set of simple oscillators, more complex motions can be assembled through *superimposition* of oscillator dynamics. Finally, the *magnet effect* occurs when separate oscillators become coupled, and a more dominant or stable oscillator pulls a less dominant oscillator into synchrony with its rhythm.

With rigid coupling of systems with identical intrinsic dynamics (as in mechanically coupled physically identical pendulums), oscillators pull into absolute coordination or entrainment, a rare phenomenon in which both the phase relationship and the frequency of oscillation match (Schmidt & Turvey, 1989). More typical scenarios exhibit a struggle between the maintenance tendency of each oscillator's intrinsic dynamics and the magnet effect, resulting in relative coordination, in which periods of synchrony alternate with periods of asynchrony. Even in cases of apparent absolute coordination, fine-grained movement dynamics often reveal a residual latent struggle between the maintenance tendency and the magnet effect.

Although von Holst's work described complex motions of oscillating fish fins, these principles have survived scale transformation to various forms of human interaction (for example, limb movements in walking, see Turvey, 1990). Interpersonal entrainment typically exhibits only relative coordination because it both lacks rigid coupling and individuals' intrinsic dynamics are never identical. For example, when individuals of different sizes walk together, their gaits might reach moments of synchrony that persist or break down, depending on their degree of coupling and intrinsic gait differences. If individuals are closer in size, they will be more likely to maintain a frequency-locked entrainment pattern because their intrinsic dynamics are more similar. If individuals physically connect by holding

hands, increased coupling might promote greater entrainment as well. In cases of relative coordination, frequencies of individual oscillators approach that of a dominant oscillator or an intermediate frequency, but any frequency match is temporary and one might observe persistent differences in phase relationships. Despite the pull to entrain to a coupled oscillator, the manifest pattern exhibits a latent influence of the original intrinsic dynamics, presumably because an external oscillator's pattern is superimposed onto an internal oscillator's pattern rather than supplanting it. Thus, an individual will never completely match the dynamics of another, but some aspects of their behavior will come into relative coordination.

The concepts of coupling, the magnet effect, and the maintenance tendency provide a ready model of the integration of internal and external forces in conversational interaction. Beek, Turvey, and Schmidt (1992) proposed that external information acts as an embedded forcing function on internal dynamics, inducing changes in the overall pattern of activity that push the activity to different values in its intrinsic range. In this way, perception can influence production through perceived external dynamics that are embedded within internal production systems, subject to degree of coupling and the maintenance tendency. Research on self-regulation of speech production, in particular the Lombard sign (Lane & Tranel, 1971) and perceptual-productive adaptation of vowel formants, speaking fundamental frequency, and consonant spectra (Houde & Jordan, 2002; Jones & Munhall, 2000; Jones & Munhall, 2003) shows that talkers can incorporate auditory feedback of their own productions to adjust subtle aspects of speech at short latencies. Perceiving the speech of other talkers might involve the same system as self-monitoring, as proposed by Gambi and Pickering (this volume). Their approach incorporates simulation through forward modeling, which is compatible with principles of entrainment (see also Fusaroli, Rczaszek-Leonardi, & Tylén, 2014).

If perception of another talker's speech yields detailed phonetic forms, such forms could influence subsequent production under circumstances, such as demands of conversational interaction, that promote coupling between talkers. At the same time, initial similarity of two talkers (e.g., whether talkers come from similar or distinct dialect backgrounds), and relative rigidity of each talker's internal dynamics will also influence degree of phonetic convergence. There is some indication that phonetic convergence during conversational interaction might be stronger for talkers from the same as opposed to different dialect regions (Kim, Horton, & Bradlow, 2011). However, there are limits to phonetic convergence which are not accounted for by similarity alone. In a study by Vallabha and Tuller (2004), talkers failed to imitate their own vowel sounds, exhibiting systematic biases in their productions that could not be accounted for by perceptual or productive processes alone.

In all studies of phonetic convergence, there were large individual differences in degree of convergence, as well as modulations by talker role and other situational factors. According to Giles' communication accommodation framework, social dominance and attractiveness influence the direction of the magnet effect pulling talkers together, but dominance is not always a straightforward consequence of talker role, and is potentially idiosyncratic to different pairs.

Finally, social dominance is irrelevant for entrainment if systems are not coupled. With looser coupling, there is likely to be less convergence, as is generally the case with informationally coupled systems, such as interacting talkers (see Schmidt & Turvey, 1989).

The notion of coupling in interpersonal coordination maps onto attention and perception, and accounts of coordination in language use will ultimately wrestle with demands of these systems. An intriguing study of coordination in rocking chairs provides direct evidence of a role for attention in rhythmic entrainment (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007). In that study, pairs of individuals sat side-by-side in rocking chairs that had the same or different natural tempos. They were either told to rock together or at their own tempo, and they either focused on a dot in front of them or on the armrest of their partner. They hypothesized that a reliance on peripheral vision would disrupt degree of coordination of their movements, however, this was only the case in the unintentional condition. When participants intended to coordinate their rocking movements, it did not matter whether they used peripheral or focal vision to couple their movements with their partner. Thus, an intention to coordinate influenced whether perceptual information that was available peripherally was ultimately used to promote interpersonal coordination. With focal attention, on the other hand, coordination emerged regardless of intention (see also Schmidt, Richardson, Arsenault, & Galantucci, 2007).

Although coupling is a critical component for entrainment, an intention to imitate is not sufficient to induce phonetic convergence in conversational interaction. Pardo *et al*. (2010) explicitly instructed one member of each pair of 12 talkers to try to imitate their partner's speech during completion of the Map Task. As in previous studies, the role of the talker who received the instruction influenced degree of phonetic convergence. Phonetic convergence was only reliable in those pairs in which Receivers had been instructed to imitate. Instructing Givers to imitate disrupted previously observed patterns of phonetic convergence. Thus, a more elaborate account of the nature of informational coupling and its relationship to intention, attention, and perception is warranted (see Schmidt, Fitzpatrick, Caron, & Mergeche, 2011; Schmidt, Morr, Fitzpatrick, & Richardson, 2012).

Conversational interaction shares many properties with other forms of interpersonal entrainment. However, unlike studies of activities such as wrist pendulum swinging, finger tapping, or chair rocking, the intrinsic dynamics of many attributes of spoken language are extremely complex and still relatively poorly understood. At a first pass, one would at least expect to find rhythmic entrainment in speech production (Wilson & Wilson, 2005), but findings of rate entrainment in speech production during conversational interaction have been inconsistent. For example, early studies in the communication accommodation literature reported moderate interlocutor correlation in speech rate (Putman & Street, 1984; Street, 1982), but more recent investigations have failed to find consistent speech rate convergence, or found significant differences in speech rate (Pardo *et al*., 2010; Pardo, Cajori Jay *et al*., 2013). In language use, as opposed to other forms of interpersonal coordination, the situation is complicated by the fact that the surface form of most

attributes is influenced by multiple linguistic and nonlinguistic (i.e., social/situational/cultural) goals simultaneously. Discovering how these multiple aims come together in acts of spoken communication across all settings is a worthwhile pursuit (see Krauss & Pardo, 2006).

Conclusion

Conversational interaction poses a challenge for psycholinguistics, both methodologically and conceptually. One of the aims of this chapter is to demonstrate that the work entailed in overcoming some of the methodological challenges can yield useful and important contributions to an understanding of speech production and perception. As Pickering and Garrod (2004) point out, language use can be thought of as a continuum from monologue to dialogue, with some occasions of monologue simulating important aspects of dialogue, and some occasions of dialogue as little more than serial monologue. Taken together, research on conversational interaction demonstrates that parity is not the only aim—sometimes talkers express themselves in ways that lead to disparity in multiple attributes of spoken communication. Although listeners expect intelligible messages, there is plenty of latitude for disparity that can enhance mutual understanding of more than words can say.

Author Note

Partial support for preparation of this chapter was provided by a grant from the National Science Foundation PAC program #BCS-1229033. The author thanks Helen Cairns, Eva Fernández, Chiara Gambi, Martin Pickering, and Robert Remez for their help in improving the narrative and supporting this work.

REFERENCES

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S.,… Weinert, R. (1991). The H.C.R.C. Map Task Corpus. *Language & Speech*, 34, 351–366.

Arantes, P., & Barbosa, P. A. (2010). Production-perception entrainment in speech rhythm. *Proceedings from Speech Prosody 2010*, Chicago.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the Intelligibility of Referring Expressions in Dialogue. *Journal of Memory and Language*, 42(1), 1–22. doi: 10.1006/jmla.1999.2667

Barr, D. J., & Keysar, B. (2002). Anchoring Comprehension in Linguistic Precedents. *Journal of Memory and Language*, 46(2), 391–418. doi: 10.1006/jmla.2001.2815

Barr, D. J., & Keysar, B. (2006). Perspective taking and the coordination of meaning in language use. In M. Traxler & M. Gernsbacher (Eds.), *Handbook of*

Psycholinguistics, *Second Edition* (pp. 901–938). Burlington, MA: Academic Press.

Beek, P. J., Turvey, M. T., & Schmidt, R. C. (1992). Autonomous and nonautonomous dynamics of coordinated rhythmic movements. *Ecological Psychology*, 4, 65–95.

Bilous, F. R., & Krauss, R. M. (1988). Dominance and accommodation in the conversational behaviours of same-and mixed-gender dyads. *Language and Communication*, 8(3/4), 183–194.

Bock. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–387.

Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, Ethnicity, & Intergroup Relations* (pp. 119–135). London: Academic Press.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75, B13–B25.

Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition*, 104(2), 163–197. doi: 10.1016/j.cognition.2006.05.006

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482–1493.

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation, Vol*. 53 (pp. 301–344). Burlington: Academic Press. doi: 10.1016/S0079-7421(10)53008-1

Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicity mention of instruments. *Cognitive Psychology*, 19, 441–472.

Byrne, D. (1971). *The Attraction Paradigm*. New York, NY: Academic Press.

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

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

Estival, D. (1985). Syntactic priming of the passive in English. *Text* 5(1), 7–22.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylen, K. (2012). Coming to Terms: Quantifying the Benefits of Linguistic Coordination. *Psychological Science*, 23, 931–939. doi: 10.1177/0956797612436816

Fusaroli, R., Raczaszek-Leonardi, J., & Tylen, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, 32, 147–157.

Fussell, S. R., & Krauss, R. M. (1989). The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology*, 25(3), 203–219.

Gallistel, C. R. (1980). Von Holst's coupled oscillators. In *The Organization of Action: A New Synthesis* (pp. 69–110). Hillsdale, NJ: Lawrence Erlbaum.

Gallistel, C. R. (1981). Précis of Gallistel's *The Organization of action: A new synthesis*. *Behavioral & Brain Sciences*, 4, 609–650.

Gallois, C., Giles, H., Jones, E., Cargiles, A. C., & Ota, H. (1995). Accommodating intercultural encounters: Elaboration and extensions. In R. Wiseman (Ed.), *Intercultural Communication Theory* (pp. 115–147). Newbruy Park, CA: Sage.

Gambi, C. & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology*, 4, 340. doi: 10.3389/fpsyg.2013.00340

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53(3), 181–215.

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, 15, 87–109.

Giles, H., Coupland, J., & Coupland, N. (1991). Chapter 1: Accommodation theory: Communication, context, and consequence. In *Contexts of Accommodation: Developments in Applied Sociolinguistics* (pp. 1–68). Cambridge: Cambridge University Press.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279.

Gregory, S. W., Dagan, K., & Webster, S. (1997). Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behavior*, 21(1), 23–43.

Gregory, S. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology*, 70(6), 1231–1240.

Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*, 34(4), 365–399. doi: 10.1007/s10936-005-6139-3

Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory & Language*, 58, 214–238. doi: 10.1016/j.jml.2007.07.003

Heldner, M., Edlund, J., & Hirschberg, J. (2010). Pitch similarity in the vicinity of backchannels. *INTERSPEECH-2010*, 3054–3057.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91–117.

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of speech, language, and hearing research*, 45(2), 295.

Jones, E., Gallois, C., Callan, V., & Barker, M. (1999). Strategies of accommodation: Development of a coding system for conversational interaction. *Journal of Language & Social Psychology*, 18, 123.

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *Journal of the Acoustical Society of America*, 108, 1246–1251.

Jones, J. A., & Munhall, K. G. (2003). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *Journal of the Acoustical Society of America*, 113(1), 532–543. doi: 10.1121/1.1529670

Keysar, B., Barr, D. J., Balin, D. J., & Brauner, J. S. (2000). Taking perspective in conversation: The use of mutual knowledge for error correction. *Psychological Science*, 11, 32–38.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, 2(1), 125–156.

Krauss, R. M. (1987). The role of the listener: Addressee influences on message formulation. *Journal of Language and Social Psychology*, 6(2), 81–98.

Krauss, R. M., & Bricker, P. D. (1967). Effects of transmission delay and access delay on the efficiency of verbal communication. *Journal of the Acoustical Society of America*, 41, 286–292.

Krauss, R. M., Bricker, P. D., Garlock, C. M., & McMahon, L. E. (1977). The role of audible and visible back-channel responses in interpersonal communication. *Journal of Personality and Social Psychology*, 9, 295–300.

Krauss, R. M., & Chiu, C. Y. (1998). Language and social behavior. In D. Gilbert, S. Fiske, & G. Lindsey (Eds.), *The Handbook of Social Psychology, 4th Edition* (pp. 41–88). Boston: McGraw-Hill.

Krauss, R. M., & Fussell, S. R. (1996). Social psychological models of interpersonal communication. In E. T. Higgins & A. Kruglanski (Eds.), *Social Psychology: Handbook of Basic Principles* (pp. 655–701). New York: Guilford Press.

Krauss, R. M., & Pardo, J. S. (2006). Speaker perception and social behavior: Bridging social psychology and speech science. In P. A. M. van Lange (Ed.), *Bridging Social Psychology: Benefits of Transdisciplinary Approaches* (pp. 273–278). Hillsdale, NJ: Lawrence Erlbaum.

Krauss, R. M., Vivekananathan, P. S., & Weinheimer, S. (1968). 'Inner speech' and 'external speech': Characteristics and communication effectiveness of socially and non-socially encoded messages. *Journal of Personality and Social Psychology*, 35, 523–529.

Krauss, R. M., & Weinheimer, S. (1964). Changes in the length of reference phrases as a function of social interaction: A preliminary study. *Psychonomic Science*, 1, 113–114.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in social interaction. *Journal of Personality and Social Psychology*, 4, 342–346.

Labov, W. (1974). Linguistic change as a form of communication. In A. Silverstein (Ed.), *Human Communication: Theoretical Explorations* (pp. 221–256). Hillsdale, NJ: Lawrence Erlbaum.

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech & Hearing Research*, 14, 677–709.

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology* 14, 78–106.

Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *INTERSPEECH-2011*, Florence, Italy, 3081–3084.

McGarva, A. R., & Warner, R. M. (2003). Attraction and social coordination: Mutual entrainment of vocal activity rhythms. *Journal of Psycholinguistic Research*, 32(3), 335–354.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, 32(5), 790–804.

Nenkova, A., Gravano, A., & Hirschberg, J. (2008). *High frequency word entrainment in spoken dialogue*. Proceedings from Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, 21(4), 337–360. doi: 10.1177/026192702237953

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2), 132–142. doi: 10.1016/j.wocn.2010.12.007

Pardo, J. S. (2000). *Imitation and coordination in spoken communication*. Doctoral dissertation, Yale University.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 119(4), 2382–2393. doi: 10.1121/1.2178720

Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology: Cognitive Science*, 4, 559.

Pardo, J. S., Cajori Jay, I., Hoshino, R., Hasbun, S. M., Sowemimo-Coker, C., & Krauss, R. M. (2013). The influence of role-switching on phonetic convergence in conversation. *Discourse Processes*, 50, 276–300. doi: 10.1080/0163853X.2013.778168

Pardo, J. S., Cajori Jay, I., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, &*

Psychophysics, 72(8), 2254–2264. doi: 10.3758/APP.72.8.2254

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, 40(1), 190–197. doi: 10.1016/j.wocn.2011.10.001

Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory & Language*, 69, 183–195. doi: 10.1016/j.jml.2013.06.002

Pardo, J. S., & Remez, R. E. (2006). The perception of speech. In M. Traxler & M. Gernsbacher (Eds.), *The handbook of psycholinguistics* (2nd edition, pp. 201–248). New York: Academic Press.

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, 79, 637–659. doi: 10.3758/s13414-016-1226-0

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02), 169–190.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3), 105–110. doi: 10.1016/j.tics.2006.12.002

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral & Brain Sciences*, 36, 49–64. doi: 10.1017/S0140525X12003238

Putman, W. B., & Street, R. L. (1984). The conception and perception of noncontent speech performance: Implications for speech-accommodation theory. *International Journal of the Sociology of Language*, 46, 97–114.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26(6), 867–891. doi: 10.1016/j.humov.2007.07.002

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25, 421–436.

Schmidt, R. C., Fitzpatrick, P., Caron, R., & Mergeche, J. (2011). Understanding social motor coordination. *Human Movement Science*, 30, 834–845. doi: 10.1016/j.humov.2010.05.014

Schmidt, R. C., Morr, S., Fitzpatrick, P., & Richardson, M. J. (2012). Measuring the dynamics of interactional synchrony. *Journal of Nonverbal Behavior*, 36, 263–279.

Schmidt, R. C., Richardson, M. J., Arsenault, C., & Galantucci, B. (2007). Visual tracking and entrainment to an environmental rhythm. *Journal of Experimental Psychology: Human Perception & Performance*, 33(4), 860–870. doi: 10.1037/0096-1523.33.4.860

Schmidt, R. C., & Turvey, M. T. (1989). Absolute coordination: An ecological perspective. In S. A. Wallace (Ed.), *Perspectives on the Coordination of Movement* (pp. 123–156).

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, 47, 1–24.

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of Discourse Processes* (pp. 123–164). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21(2), 211–232.

Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society*, 5, 1–23.

Shepard, C. A., Giles, H., & Le Poire, B. A. (2001). Communication Accommodation Theory. In W. P. Robinson & H. GIles (Eds.), *The New Handbook of Language & Social Psychology* (pp. 33–56). New York: Wiley.

Street, R. L. (1982). Evaluation of noncontent speech accommodation. *Language & Communication*, 2, 13–31.
Turvey, M. T. (1990). Coordination. *American Psychologist*, 45(8), 938–953.
Vallabha, G. K., & Tuller, B. (2004). Perceptuomotor bias in the imitation of steady-state vowels. *Journal of the Acoustical Society of America*, 116(2), 1184–1197. doi: 10.1121/1.1764832
Weiner, E. J. & Labov, W. (1983). Constraints on the agentless passive. *Journal of Linguistics* 19, 29–58.
Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory & Language*, 31, 183–194.
Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, 12(6), 957–968.