6-1-2019

The Montclair Map Task: Balance, Efficacy, and Efficiency in Conversational Interaction

Jennifer Pardo
Montclair State University, pardoj@montclair.edu

Adelya Urmanche
Montclair State University

Hannah Gash
Montclair State University

Jaclyn Wiener
Montclair State University

Nicholas Mason
Montclair State University

See next page for additional authors

Article

Language and Speech

Language *and* Speech
2019, Vol. 62(2) 378–398

DOI: 10.1177/0023830918775435
journals.sagepub.com/home/las
SAGE

The Montclair Map Task: Balance, Efficacy, and Efficiency in Conversational Interaction

Jennifer S Pardo, Adelya Urmanche, Hannah Gash, Jaclyn Wiener, Nicholas Mason, Sherilyn Wilman, Keagan Francis, and Alexa Decker
Department of Psychology, Montclair State University, USA

Abstract

This paper introduces a conversational speech corpus collected during the completion of a map-matching task that is available for research purposes via the Montclair State University Digital Commons Data Repository. The Montclair Map Task is a new, role-neutral conversational task that involves paired iconic maps with labeled landmarks and a path drawn from a start point, around various landmarks, to a finish mark. One advantage of this task-oriented corpus is the ability to derive independent objective measures of task performance for both members of a conversational pair that can be related to aspects of communicative style. A total of 96 native English speakers completed the task in 16 same-sex female, 16 same-sex male, and 16 mixed-sex pairings. Conversations averaged 32 minutes in duration, yielding approximately 217,000 words. The transcription protocol delineates events such as speaking turns, inter-turn intervals, landmark phrases, fillers, pauses, overlaps, and backchannels, making this corpus a useful tool for investigating dynamics of conversational interaction. Analyses of communication efficacy and efficiency reveal that male pairs of talkers were less efficient than female and mixed-sex pairs with respect to partner map-matching task performance.

Keywords

Conversational interaction, conversation corpus, communication efficiency

1 Introduction

Conversations occur in a variety of settings, from informal chats to highly constrained formalized interactions with specialized vocabulary. This paper introduces a new conversational speech corpus collected in a setting that falls between these extremes, during the completion of a moderately

Corresponding author:
Jennifer S. Pardo, Psychology Department, PhD, Montclair State University, 1 Normal Avenue, Montclair, NJ 07043, USA.
Email: pardoj@montclair.edu

constrained map-matching task. The Montclair Map Task (MMT) is a modified, role-neutral version of the HCRC Map Task (Anderson et al., 1991). Like the original HCRC Map Task, the MMT involves paired iconic maps with labeled landmarks and a path drawn from a start point, around various landmarks, to a finish mark. However, the MMT neutralizes the role difference present in the original version and focuses attention on the landmark labels during completion of the task in order to promote interactions with more balanced contributions between members of a pair. One advantage of this task-oriented corpus is the ability to derive independent objective measures of task performance that can be related to aspects of communicative style. In addition, the corpus supports analyses of sex differences in communicative style by including equal numbers of male and female talkers in same- and mixed-sex pairings. Thus, the current paper describes collection and transcription of the MMT corpus, presents descriptive analyses of corpus events, and offers an analysis of communicative balance, efficacy, and efficiency in task-oriented conversational interaction. To promote future research on the dynamics of conversational interaction, the corpus recordings and associated annotation files are available for research purposes at an online digital repository (Pardo et al., 2018).

One of the aims of collecting the MMT corpus was to broaden investigations of phonetic variation and convergence in conversational interaction (Pardo, 2006; Pardo, Cajori Jay, & Krauss, 2010; Pardo, Cajori Jay, Hoshino, Hasbun, Sowemimo-Coker, & Krauss, 2013). As reported in Pardo et al. (2018), between-talker phonetic convergence occurred in the MMT conversations and was characteristically subtle and variable across talkers, but the corpus can also inform investigations of more general aspects of conversational dynamics. The original HCRC Map Task was a useful tool for investigating phonetic convergence because it includes a set of pre-determined, phonologically diverse landmark label phrases, which are maintained in the current version of the task. In its original form, the HCRC Map Task involved a role distinction between an instruction giver, whose map contained a path, and a receiver, who was tasked with drawing the giver's path on their map. The role distinction was useful for investigating the influence of social role, but this aspect of the HCRC task's design has three unfortunate consequences. First, the task yields unbalanced contributions from members of a pair of talkers, such that givers produce more utterances overall than receivers. An analysis of data from Pardo (2006), which used the HCRC Map Task, reveals that givers produced 2.7 times as many words as receivers, with turn durations that were 3.5 times longer. Second, the task also evokes unbalanced utterances of landmark label phrases between members of a pair—instruction givers in Pardo (2006) produced twice as many landmark labels as receivers. Finally, the task provides measures of performance for only one member of a pair, the instruction receiver.

In order to alleviate these concerns and neutralize effects of nominal role, the MMT modification involves reconciling differences between the maps in the composition of the landmarks. That is, the maps were altered so that each pair member's map contains five shared and five unique landmarks, and the goal of the task is for both interlocutors to discover the unshared landmarks that appear on their partner's map and to draw markers for them on their own map. Because both members of a pair have information to share and a task to complete, the MMT also provides a measure of task performance for both interlocutors. Thus, in addition to investigating phonetic variation among landmark label phrases, aspects of each talker's communicative style can be analyzed in relation to their partner's performance on the map task.

In comparison with other conversational speech corpora, the MMT is unique with respect to its ability to yield an independent objective measure of communicative efficacy. Most currently available transcribed conversational speech corpora involve unconstrained conversations on everyday topics or an interview format (e.g., Buckeye corpus: Pitt, Johnson, Hume, Kiesling, & Raymond, 2005; CALLHOME corpus: Canavan, Graff, & Zipperlen, 1997; Fisher corpus: Cieri, Miller, &

Walker, 2004; Philadelphia Neighborhoods corpus: Labov, Rosenfelder, & Fruehwald, 2013; Switchboard corpus: Godfrey & Holliman, 1993). Apart from the HCRC Map Task, the Diapix task used in the Wildcat corpus is most similar to the MMT (Van Engen, Baese-Berk, Baker, Choi, Kim, & Bradlow, 2010; see also Walking Around corpus: Brennan, Schuhmann, & Batres, 2013). In the Diapix task, interlocutors discuss images of scenes with 10 differences between them, and must spot the differences between the images. The use of shared and unshared landmarks in the MMT was inspired by this aspect of the Diapix task. However, performance on the Diapix task in the Wildcat corpus was near perfect, yielding insufficient variation for assessing effects of communicative style on efficacy. Although identification of missing landmarks was likewise very strong in the MMT, precision of landmark placement on the maps varied sufficiently to assess more subtle aspects of communicative efficacy.

Another useful feature of the current corpus is a design with a balanced set of same- and mixed-sex pairings—talkers interacted in 16 same-sex female, 16 same-sex male, and 16 mixed-sex pairings. Reports of talker sex effects on various aspects of language use abound in the literature, often with conflicting results that warrant further investigation. The complexity of sex effects on conversational interaction is well-documented in a comprehensive meta-analysis reported by Leaper and Ayres (2007). They examined 63 published studies of sex effects on talkativeness that used quantitative analyses from 1968 to 2004, yielding measures of Cohen's *d* effect size for 70 independent samples comprising a total of 4385 participants. Across the set of studies, there was a small but significant tendency for men to talk more than women overall (Cohen's *d* averaged −0.14; coding was positive if females were more talkative), but this pattern was highly variable, depending on dyad composition, a talker's relationship with an interlocutor, and type of activity. In studies that examined mixed-sex dyads, men were more talkative than women, regardless of topic, but only when their female interlocutors were either strangers (−0.38) or spouses (−0.46)—if their female interlocutors were friends, then men were less talkative than women (0.38). In contrast, studies that examined same-sex pairings found that female pairs were more talkative than male pairs, but only when interlocutors were strangers in an unstructured (0.82) or assigned decision-making setting (1.02)—in a setting with strangers discussing an assigned general topic, male pairs were more talkative than female pairs (−1.50).

Due to the complexity of sex effects in the literature, a comprehensive conversational speech corpus must include male and female talkers interacting in same- and mixed-sex pairings. Given that talkers in the current study were strangers interacting in a relatively constrained setting, a prediction consistent with previous findings is that males will produce more words overall than females, particularly in mixed-sex pairings. In same-sex pairings, females should produce more words than males. Moreover, because the current corpus permits a measure of conversational efficacy with respect to map task performance, it will also be possible to examine whether observed sex differences in talkativeness bear any relation to communicative efficacy. For example, when a male talker in a mixed-sex pairing produces more words, does the task performance of their female partner benefit?

The MMT corpus comprises recordings of 96 native English speakers (48 female) from a mostly homogenous dialect region who completed the new map task with six pairs of maps. To permit assessment of effects of talker sex, talkers were randomly assigned to 16 same-sex female, 16 same-sex male, and 16 mixed-sex pairs. Conversations ranged from 16 to 62 minutes in duration to task completion, averaging 32 minutes. Research assistants segmented and orthographically transcribed the two-channel digital recording files to delineate speaking turns, inter-turn intervals, landmark label phrases, pauses, fillers, backchannels, and overlapping speech events. The quality of the audio recordings, with two-channel talker separation in a soundproof booth, permits detailed analysis of acoustic–phonetic attributes. In particular, the set of 79 landmark label phrases was designed to elicit a full set of English vowel sounds as well as to provide opportunities to observe various phonological phenomena such as regressive place assimilation (e.g., *green bay*) and a variety of dialect markers in American English. For example, the phrases *greasy wash water* and *oily rag* were drawn from the dialect calibration sentences used in the TIMIT corpus (Fisher, Doddington, & Goudie-Marshall, 1986), and additional phrases known to distinguish dialects included *milk bar, Caribbean palm, garage, huge nuclear plant, north square*, and *orange car* (Labov, Ash, & Boberg, 2006). Descriptive analyses of the corpus focus on relative balance in talker contributions within each pair, counts of corpus events, and efficacy in terms of map task performance. To provide a demonstration of the unique contributions of this corpus, an analysis of conversational efficiency relates performance on the map task to total number of words and percentage landmark phrases in particular.

2 Methods

2.1 Participants

In total, 96 native English speakers (48 female) with normal hearing and speech from the Montclair State University community participated in paired conversational interactions. Each participant provided IRB-approved informed consent and received US$20 compensation for their time. As part of the informed consent procedure, each participant explicitly agreed to the sharing of their de-identified data in other studies. Recordings took place in March–May of 2013 for the first 24 pairs and in May of 2014 for the last 24 pairs.

A demographic questionnaire asked participants for information about their residential history, ethnicity, race, family income level, and language background, as shown in Appendix 1 (one participant did not complete a demographic questionnaire due to experimenter error—a male from same-sex male pair 34). Participants ranged in age from 18 to 38 years old, with an average age of 21 years at the time of recording (*SD* = 2.8 years). When asked to list all the cities they had lived in from birth to the time of recording, 71 participants reported being born in New Jersey. Of the remainder, 11 had been born in New York, two in Pennsylvania, one in Illinois, one in Minnesota, one in Colorado, one in Texas, two in the Dominican Republic, one in Puerto Rico, one in Jamaica, one in Poland, one in Kyrgyzstan, and one in Sri Lanka. Most participants reported having lived in two or more places ($n = 62$), and nearly all participants had resided in New Jersey for at least three years prior to the time of recording (one was still living in Pennsylvania, and two had lived in New Jersey for only one year prior). Of the seven participants born outside the United States, all had resided in New Jersey for five or more years prior to recording. Thus, nearly 90% of participants had been born in and lived most of their lives in the New Jersey/New York/Pennsylvania area.

With respect to ethnicity, 20 participants identified as Hispanic/Latino when asked to select either Hispanic/Latino or Not Hispanic Latino (two participants left the question blank). With respect to race, 61 participants identified as White, 15 as Black, four as Asian, two as American Indian or Alaska Native, one as Black/White, one as Native Hawaiian/White, one as Other, and one wrote Latina (eight participants left the question blank). Half of the participants reported an annual family income of $70,000 per year or less (eight participants left the question blank).

A final question asked participants about their language background. The question prompted participants to list any languages they knew in addition to English, to indicate the year they had started learning the language, and to describe their fluency level either as fluent, at a basic conversational level, or used in a school setting only. In their responses, 26 participants reported being fluent in a second language, and 19 participants left the item blank. It is likely that some of those reporting fluency would be considered bilingual, since many listed their birth year as the year that they began learning the non-English language.

2.2 Materials

The conversational task involved six pairs of iconic maps that had paths drawn around labeled landmarks. Each member of a pair received a packet of six maps printed on 8.5″ × 11″ sheets of paper that corresponded with their partner's map packet. Each map in a corresponding pair contained five landmarks that also appeared on the partner's map (shared) and five landmarks that were unique to that map. In addition, both maps in a pair contained an identical path drawn on the map, from a starting point, around various landmarks, to a finish. The set of landmarks and the shape of the map path varied between map pairs. Across the full set of maps, there were 79 unique landmark label phrases, with some appearing on more than one map. Appendix 2 displays a sample pair of maps, and Appendix 3 lists the full set of landmark label phrases printed on the maps.

2.3 Procedures

Talkers formed randomly assigned pairs: 16 same-sex female, 16 same-sex male, and 16 mixed-sex. None of the talkers had been acquainted with their partners prior to the recording session. All recordings took place in an Acoustic Systems (IAC Acoustics, North Aurora, IL) sound proof booth. Each talker sat in the booth at a small table that was separated from their partner's table by a room divider that prevented the talkers from seeing each other, and used a pencil to draw on the maps as they completed the conversational task. Participants wore AKG (AKG Acoustics, Vienna, Austria) head-mounted microphones connected to an Apple (Apple Inc, Cupertino, CA) Macintosh computer situated outside the booth. SoundStudio (Felt Tip Inc, New York, NY) software recorded each talker onto an individual time-aligned channel in a two-channel audio file. Recordings were digitized at 44.1 kHz and saved as 16-bit files in .wav format.

2.3.1 Conversational task. The Montclair Map Task is a modified version of the HCRC Map Task (Anderson et al., 1991). The MMT involves paired iconic maps with labeled landmarks, and the goal of the task is for talkers to reconcile differences across the maps in the composition and location of the landmarks. This procedure evokes natural conversational interaction while ensuring that a set of pre-determined phrases (landmarks) will be repeated between talkers. Each map in a pair contains five shared and five unique landmarks. Instructions inform both talkers in a pair that they must discover the identity and location of the five landmarks that only appear on their partner's maps (the unique ones) and draw labeled markers indicating the location of the missing landmarks on their own maps as accurately as possible, making this a map-matching task. To provide shared reference frames for the task, both maps in each pair also have an identical path drawn on them from a start, around various landmarks, to a finish.

2.3.2 Transcription protocol. Trained research assistants completed segmentation and orthographic transcriptions of all 48 conversations using annotation procedures in Praat software (freeware available at www.praat.org). The Quick Rich Transcription (QRTR) Specification for English Broadcast Data protocol developed by the Linguistic Data Consortium (www.ldc.upenn.edu) inspired the transcription routine for these data, which involved segmentation of the audio files, and transcription and labeling of the content. Each audio file comprised two time-aligned channels, one for each talker. Annotators segmented the two-channel audio files to delineate epochs for *map number* (1–6), *inter-map intervals* (time from the end of discussing one map pair to the beginning of discussing another map pair), *speaking turns* (utterance by a single talker accomplishing a communicative goal; Sacks, Schegloff, & Jefferson, 1974), *inter-turn intervals* (time from the end of one talker's turn to the beginning of their partner's turn), *backchannels* (utterances produced by a

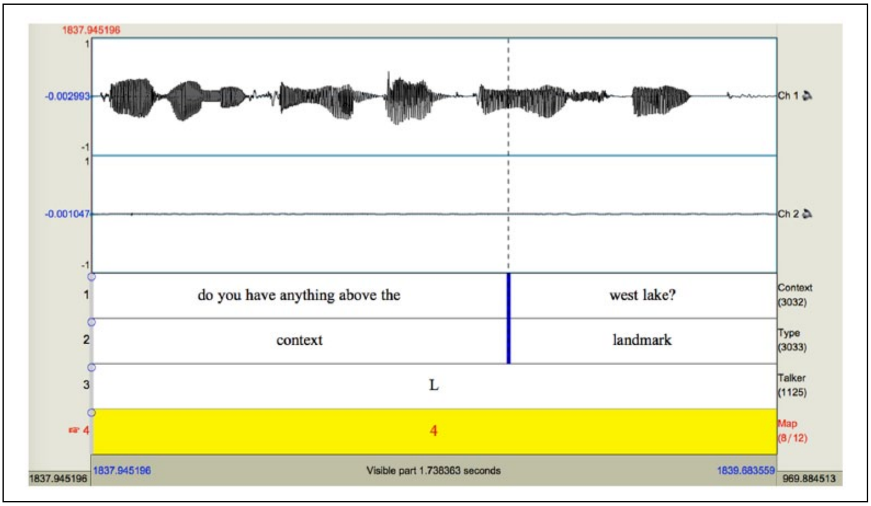

Figure 1. Sample Praat annotation in the Montclair Map Task corpus. This illustration shows a sample section of a two-channel recording from a single pair of talkers, as annotated in Praat. The sections in the top panel show the time-aligned waveforms in each channel (the channel 2 talker was silent during this sample). The bottom panel shows the four tiers of annotation: 1. the transcribed Content; 2. the Type label tier; 3. the Talker code; and 4. the Map Order code. In this sample, the channel 1 talker says *do you have anything above the west lake*. In accordance with the annotation protocol, the *west lake* landmark label was segmented into a single epoch labeled *landmark*. The remainder of the utterance was segmented into a separate phrasal epoch and labeled *context*.

partner during the main talker's turn that do not initiate a new turn, e.g., "uh-huh," "okay," etc.), *overlapping speech* (events in which both talkers speak simultaneously), *pauses* (within-turn silent intervals > 100 ms), *fillers* (pause fillers such as "uh," "um," etc.), *non-speech sounds* (laughter, lip-smacking, etc.), and *landmark label phrases*.

Note that for speech content, words and phonemes were not segmented individually—only complete landmark label phrases were delineated from non-landmark utterances. Annotators segmented words produced within a turn that were not part of a landmark label phrase into longer epochs labeled *context*. For example, Figure 1 displays a sample Praat annotation file in which a talker said "Do you have anything above the west lake." In this case, the utterance was segmented into two epochs, "Do you have anything above the" and "west lake." The first segment was transcribed verbatim and labeled *context*, and the second segment was transcribed and labeled *landmark*. Praat textgrid files stored the data from segmentation, transcription, and labeling, and conversational analyses used data extracted from these files.

3 Results and discussion

One of the main goals of creating the MMT was to elicit more balanced utterances from talkers within each pair. As noted in the Introduction, givers in the HCRC Map Task recordings from Pardo (2006) produced 2.7 times as many words as receivers and twice as many landmark label phrase utterances. An examination of within-pair differences reveals that utterance quantities more closely matched in the MMT, averaging 30% (*SEM* = 4%) more words total and 29% (*SEM* = 4%) more landmark words produced by the more talkative members of each pair, and there were no sex differences in these comparisons. Thus, the MMT was successful in eliciting relatively more balanced contributions within each pair of talkers across all pair types (same-sex female, same-sex male, mixed-sex). The following analyses examine accuracy in Map Task performance, report counts of Map Task events, and offer an analysis of communicative efficiency by relating measures of talkativeness to task performance.

3.1 Map task accuracy

From the perspective of a participant, the goal of the task was to draw markers indicating the locations of five missing landmarks on each of six maps by engaging in conversation with their partner, while their partner did the same for their maps. Accuracy in locating missing landmarks was very high—failures to include all five landmarks occurred on only six out of 576 maps and only one map contained an extra landmark. A finer-grained analysis of map-matching performance indicated that although very few maps had missing landmarks, precision in location of landmark placement varied. In order to measure precision in landmark location, two independent raters examined each landmark's location and assigned its placement to one of five categories: *Correct* location, *Adjacent* location on the same side of the path, *Nearby* location on the wrong side of the path, *Distant* location (more than two inches away in any direction without regard to path), and *Missing*. Then, each rater allocated 20 points for each correctly placed landmark, 15 points for adjacent, 10 points for nearby, 5 points for distant, and 0 points for missing landmarks. Thus, each map could receive up to 100 points if a rater judged all five landmarks to be correctly placed. Inter-rater reliability for this measure was high ($r = 0.84$), and final precision scores comprised averages across the two raters' scores.

Using the landmark location-based measure of precision in completion of the map task, performance levels averaged 90 points ($SD = 8.02$; out of 100 total points possible). Examining the data more closely, accuracy increased over the course of conversations from 86 points on Map 1 to 92 points on Map 6. An analysis of variance including map order and talker sex confirmed an increase in performance across maps, $F(5, 470) = 11.34$, $p < .0001$, $\eta^2 = 0.047$, and revealed no significant differences across female and male participants in task performance and no interaction (ps > 0.6). The amount of time spent completing the entire map task session averaged 32 minutes ($SD = 10.68$ minutes), ranging from 16 to 62 minutes, and was weakly correlated with average map task performance, $r(46) = 0.28$, $p = 0.05$. Female pairs took slightly longer to complete the task on average (35 minutes, $SD = 10.11$) than male pairs (33 minutes, $SD = 10.79$) and mixed-sex pairs (31 minutes, $SD = 11.52$), but pair sex differences in time to complete the task were not significant. In general, talkers communicated effectively to complete the task, with variability in precision of landmark label placement that was weakly related to the amount of time to complete the task, and that did not differ by talker or pair sex. The variability in placement of landmark label location can be used as a measure of relative efficacy in communication.

3.2 Corpus event counts

Table 1 lists counts of turns, words, syllables, landmarks, pauses, fillers, overlaps, and backchannels in the corpus. To assess whether the current corpus aligns with patterns of sex differences reported in the literature, the table also splits counts according to talker sex and pair sex. The last two columns split the data within mixed-sex pairs alone by talker sex. Counts of words and syllables include all context words, landmarks, and fillers produced by the main talker in a turn, but not utterances produced by a partner as overlapping speech or as backchannels during the main talker's turn. Counts of pauses reflect all within-turn silent intervals greater than 100 ms in duration, including any that were adjacent to fillers or backchannels, but excluding silent intervals at turn-exchanges.

Table 1. Counts of corpus events across talker sex and pair sex.

	All pairs	Talker sex		Pair sex			Mixed-sex pairs	
		Female	Male	Female	Male	Mixed	Female	Male
Turns	20,027	9,791	10,236	6,629	7,068	6,330	*3,162*	*3,168*
Words	216,468	99,992	116,476	69,566	78,199	68,703	30,426	38,277
Syllables	260,690	120,293	140,397	83,487	94,537	82,666	36,806	45,860
Landmarks	13,868	6,392	7,476	4,348	5,148	4,372	2,044	2,328
Pauses	22,648	*11,460*	*11,188*	8,054	7,765	6,829	*3,406*	*3,423*
Fillers	4,881	2,228	2,653	1,502	*1,696*	*1,683*	726	957
Overlaps	5,472	2,256	3,216	1,453	2,196	1,823	*896*	*927*
Backchannels	7,207	3,719	3,488	*2,397*	*2,391*	*2,419*	1,322	1,097
Landmarks	64	64	64	63	66	64	67	61
Pauses	105	115	96	116	99	99	112	89
Fillers	23	22	23	22	22	24	24	25
Overlaps	25	23	28	21	28	27	23	30
Backchannels	33	37	30	34	31	35	35	36

The top section of the table lists summed counts by interaction type. The bottom section of the table lists counts per 1000 words. Word and syllable counts include landmarks and fillers. Counts of landmarks and fillers do not include those produced as overlapping speech. Counts of pauses do not include intervals at turn-exchanges. Overlaps count utterances of a talker during their partner's turn (excluding overlapped backchannels). Comparisons across talker sex in the top section are significant (χ-square *p*s < 0.03), except for those in italic font.

The overlaps category counts all incidents of context speech, landmarks, and fillers that a talker produced during their partner's turns that overlapped with the partner's speech, and those utterances are not included in counts for other categories (i.e., a talker's overlapping utterances of landmarks and fillers are counted together as overlaps, and not with other counts for landmarks and fillers). Note that overlapping backchannels are not included in counts for overlaps. Counts of backchannels comprise utterances produced by a talker during their partner's turns that were meant to encourage the talker to continue their turn, and 33% of these overlapped with a partner's speech. Most backchannels were produced completely during a partner's pause, so they overlapped with a partner's turn, but not with their speech.

The top section of Table 1 lists summed counts for all interactions, and the bottom section lists counts of events per 1000 words. All comparisons of event counts by talker sex, pair sex, and within mixed-sex pairs listed in the top section of the table were significant in χ-squared tests except that pause counts did not differ by talker sex overall; backchannel counts did not differ across pair sex categories; and turn, pause, and overlap counts did not differ across mixed-sex pairs (significant χ-squared values ranged from 3.19 to 1550.40, with all *p*s < 0.03; non-significant comparisons are italicized in the table). The bottom section of the table lists event counts relative to the number of words produced, thereby taking into account opportunities to observe events. Rates for overlaps and backchannels in mixed-sex pairs were adjusted according to the number of words produced by the opposite-sex partners because these were produced during their turns.

The data listed in the Talker sex columns in Table 1 reveal that male talkers produced more instances of all event types than female talkers, except that females paused slightly more often than males; marginal $\chi^2(1) = 3.27$, $p = 0.07$. Likewise, across the Pair sex columns, same-sex male pairs produced more instances of all event types than female pairs except pauses (females produced more) and backchannels (ns), and more instances than mixed-sex pairs of all event types except fillers and backchannels (ns). As shown in the comparisons for pair sex, female pairs paused more often than male and mixed-sex pairs (8054 > 7765); $\chi^2(1) = 5.28$, $p = 0.02$. Within mixed-sex pairs, males produced more words, landmarks, and fillers than females. Whereas females paused more often than males, but only in same-sex pairings, males produced more fillers than females, both in same and mixed-sex pairings. Incidents of overlapping speech were more frequent among males than females in same-sex pairs, but overlap counts did not differ between talkers in mixed-sex pairs. In contrast, females produced more backchannels than males in mixed-sex pairings (1322 > 1097); $\chi^2(1) = 20.93$, $p < 0.0001$.

With respect to gross counts, male talkers produced more words than female talkers, in both same-sex (12% more) and mixed-sex (26% more) pairings, including producing more landmarks and fillers than females, and males used more turn-exchanges than females to do so. Males also produced more incidents of overlapping speech than females, except in mixed-sex pairs. In mixed-sex pairings, females produced more backchannels during male partners' turns than vice versa. These patterns are consistent with the finding in the meta-analysis reported in Leaper and Ayres (2007) that males were more talkative than females when conversing with strangers in mixed-sex pairs. However, male talkers in the current corpus were also more talkative than females in same-sex pairings, which further contributes to the variability in findings reported in the meta-analysis.

A closer examination of some of the individual studies analyzed by Leaper and Ayres reveals that the pattern of sex effects in same-sex pairs is generally more variable than that of mixed-sex pairs. For example, Bilous and Krauss (1988) examined a single set of unacquainted talkers interacting in an assigned decision-making setting in both same- and mixed-sex pairings (task order was counterbalanced; N.B., the pair composition for the second condition of this study was mislabeled as same gender in Leaper & Ayres 2007). In mixed-sex pairings, there was a weak nonsignificant tendency for males to talk more than their female interlocutors (Cohen's $d = -0.18$), and in same-sex pairings, females produced 22% more words than males (Cohen's $d = 2.04$). Among five other studies that examined sex effects in both same- and mixed-sex groups, all found that males in mixed-sex pairings were more talkative than females, but sex effects for same-sex pairings varied (Cashdan, 1998; McMillan, Clifton, McGrath, & Gale, 1977; Mulac, 1989; Simkins-Bullock & Wildman, 1991; Wood, 1966). For example, Wood (1966) found that males were more talkative than females in same-sex pairings (Cohen's $d = -2.14$). However, Wood measured talkativeness in terms of mean utterance length, while the finding from Bilous and Krauss (1988) relied on a measure of total number of words. Moreover, the picture description task used by Wood was not a fully-interactive conversation—the confederate targets of messages provided highly constrained or no feedback. In a second set of measures not examined by Leaper and Ayres, Bilous and Krauss also examined mean utterance length, finding that while males in same-sex pairs produced fewer words overall, the length of each utterance was longer than that of females. Despite differences in task and measure, sex differences in the current corpus align more closely with the pattern of greater male talkativeness exemplified in Woods than in Bilous and Krauss. Thus, the current observations in a task-oriented setting reinforce a general pattern of greater talkativeness with female strangers on the part of male talkers and, in this case, the pattern was consistent across same- and mixed-sex pairings. The next set of analyses further explores these stylistic differences in relation to efficiency in communication.

3.3 Communicative efficiency

Efficiency in communication is a function of time and/or effort in relation to efficacy. A talker who spends less time and/or exerts less effort cannot be considered an efficient communicator if their partner's performance suffers as a result. Thus, an appraisal of conversation efficiency in this setting must take into account a consideration of performance on the map-matching task. Recall that when averaging across pairs, there were no significant group differences by talker sex or pair sex on either the amount of time to complete the map task or task performance, and that task completion time was only weakly correlated with task performance. However, male talkers produced more words than same-sex female pairs of talkers, mixed-sex pairs overall, and female partners in mixed-sex pairs, in roughly the same amount of time. An analysis of speaking rates reveals that males accomplished this feat by speaking faster than females. As described below, although male talkers produced more words in the same amount of time as female talkers, their use of more words was not always associated with improvements in their partners' performance levels. These patterns indicate that differences in communication style of male talkers made them somewhat less efficient than female talkers.

The analysis of speaking rates used two measures, one that includes pauses (speaking rate) and one that reflects more precisely the rate of speech articulation while speaking (articulation rate). To calculate speaking rates, the number of words in each speaking turn was divided by the number of seconds in the turn (including within-turn pauses, but not inter-turn intervals), yielding a measure of words per second (wps). An analogous treatment yielded articulation rate scores, by dividing the number of words by the number of seconds articulating speech during the turn, which entailed subtracting pause durations from turn duration. In order to eliminate outliers, the following analyses are based on data between 5% and 95% of the full distribution. On average, speaking rates were higher for males than females by about 0.33 wps; 2.68 (0.33) > 2.35 (0.31) wps; $F(1, 94) = 34.55$, $p < 0.0001$, $\eta^2 = 0.21$. Across the six maps in the task, there was a small but significant tendency to reduce speaking rates by about 0.10 wps that was consistent across talker sex (males from 2.73 to 2.64, and females from 2.41 to 2.31); $F(5, 470) = 3.67$, $p < 0.003$, $\eta^2 = 0.01$ (interaction between talker sex and map ns). Articulation rates exhibited an analogous difference of about 0.35 wps, indicating that males produced more words than females even when taking into consideration the amount of time spent pausing; male 3.01 (0.32) > 2.66 (0.33) female; $F(1, 94) = 40.99$, $p < 0.0001$, $\eta^2 = 0.23$.

This preliminary analysis of group differences indicates that male talkers exhibited greater levels of speech productivity than females in the same amount of time, which was associated with roughly equivalent partner task performance levels on average. This difference in communication style made males less efficient communicators than females in this task with respect to partner task performance, but not amount of time. Although there were no significant differences in precision of landmark placement between males and females when averaged across pairs, relative efficiency in communication should also be assessed in more detailed analyses across individual talker pairs. That is, apart from group differences in baseline speech productivity, when male talkers produce more words, their partners' performance levels on the map task should improve to a analogous degree as when female talkers produce more words. Finally, other aspects of communicative style, such as relative use of landmark label phrases, could also influence some of these patterns in conversation efficiency.

3.3.1 Number of words and map task performance. In order to examine communicative efficiency within individual pairings, this set of analyses correlated each talker's total number of words in the task with map task performance by their partner. The results of these pair-wise correlation analyses split by same- and mixed-sex pair types appear in Table 2. The top section of the table lists correlation coefficients for relationships between talker data (number of words/landmarks) and map task landmark precision scores of partners (96 pairs of observations split by pair and talker sex), and the bottom section lists means and standard deviations for the by-partner and by-talker data used in the correlation analyses. The left-most set of columns display data split by pair sex, and the right-most set of columns display data split within mixed-sex

Table 2. Correlations between talker words and task performance of partner.

	Male pairs		Female pairs		Mixed pairs		Males to females		Females to males	
Correlation with task										
r number of words	−0.33*		0.60**		0.48**		0.56**		*0.37*	
r landmark words	−0.32*		0.31**		0.33**		0.55**		*0.04*	
r percentage landmarks	*0.06*		−0.76**		−0.68**		−0.75**		−0.65**	
Averages										
Partner task performance	90	(8.38)	90	(9.1)	90	(6.64)	90	(6.62)	91	(6.85)
Number of words	2444	(898)	2174	(878)	2147	(1071)	2392	(1277)	1,902	(783)
Number of landmark words	283	(107)	238	(55)	239	(74)	252	(82)	226	(66)
Percentage landmark words	11.7	(1.75)	12.2	(3.85)	12.4	(3.50)	11.8	(3.09)	13.0	(3.88)

**$p < 0.05$; *$0.056 < p < 0.08$.

pairs of talkers (*Males to Females* refers to data during male talkers' turns, while *Females to Males* refers to data during female talkers' turns). In order to confirm that the effects reported below were not driven by apparent outliers in map task performance, all correlation analyses were repeated on a subset of the data that excluded nine data points corresponding with map task performance less than 80 points (see Figures 2 and 3 below for the full dataset). The resulting correlation coefficients followed the same overall pattern as that in the full dataset, but were reduced and in some cases nonsignificant due to the resulting lack of power (data from the subset with performance > 80 points appears in Appendix 4). Because the overall patterns were similar with these low performers removed, all analyses reported below examined the full dataset to preserve power.

As displayed in Table 2, female pairs showed the strongest positive correlation between talker number of words and partner task performance (first row of table), $r(30) = 0.60$, $p < 0.0003$, followed by mixed-sex pairs, $r(30) = 0.48$, $p < 0.008$. In contrast, male pairs showed a marginally-significant negative correlation, $r(30) = -0.33$, $p = 0.06$, indicating that production of more words by male talkers was associated with reduced precision in landmark placement by male partners. Within mixed-sex pairs, however, the number of words produced by male talkers was associated with increased map task performance by their female partners, $r(14) = 0.56$, $p < 0.03$, and an analogous correlation between female talker words and performance by male partners was weaker and not significant, $r(14) = 0.37$, $p = 0.16$ (but note that this correlation relies on half as many pairs as the comparable coefficient found among same-sex male pairs).

To assess this pattern of correlations, linear mixed-effects modeling analyses examined whether number of words produced by a talker and pair sex predicted partner map task performance. In this model and an analogous model for percentage landmark labels reported below, number of words (and percentage landmarks) was *z*-scale normalized prior to entering the model, pair sex treated male as the base level (vs. female and mixed), and the model included random intercepts by talker (lme4 version 1.1-12, lmerTest version 2.0-33, R version 3.3.2; Baayen, 2008; Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013; Bates, Mächler, Bolker, & Walker, 2015; Bates et al., 2016; Kuznetsova, Brockhoff, & Christensen, 2016; Quené & van den Bergh, 2008; R Development Core Team, 2016). The modeling analysis confirmed that the relationship between talker number of words and partner task performance for male pairs differed from female pairs, $ß_{female} = 2.74$ (1.25), $t = -2.199$, $p = 0.028$, but the comparison with mixed-sex pairs was not significant, $ß_{mixed} = 0.82$ (1.22), $t = 0.67$, $p = 0.5$.

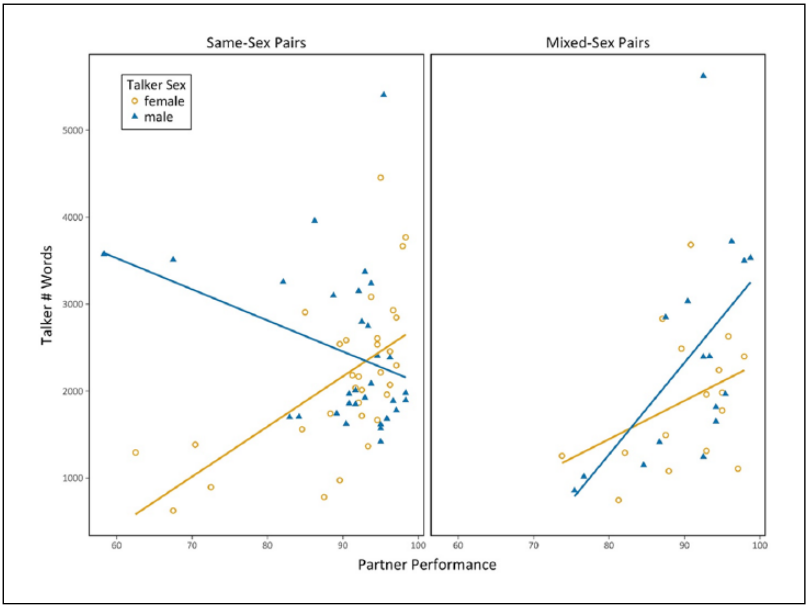

Figure 2. Correlations between a talker's number of words and their partner's task performance. The left panel displays data from same-sex pairs, and the right panel displays data from mixed- sex pairs. Female talkers are plotted with open orange circles and male talkers are plotted with filled blue triangles. Lines represent linear fits. The relationship for males in same-sex pairs differed from other pair types.

Figure 2 displays scatter plots for the relationship between the number of words produced by a talker and their partner's performance on the map task across same and mixed-sex pairs of talkers. According to this analysis, the strongest positive relationships between talker number of words and partner task performance occurred in pairs with female as opposed to male partners, regardless of the sex of the talker. Moreover, the relationship between number of words and performance of male partners was positive only when they interacted with female talkers (mixed-sex female talkers; but note that the relationship was weak and not significant). It is important to keep in mind that average partner task performance was equivalent across pair types (see data at the bottom of Table 2). Thus, roughly equivalent performance levels on the part of female partners was achieved with 9% fewer words from female talkers than male talkers (2174 vs. 2392 on average). Nonetheless, more words from talkers mostly helped female partners' performance. In contrast, male partner performance only benefitted from more words when interacting with female talkers, who used 22% fewer words than male talkers with male partners (1902 vs. 2444 on average). Because increases in numbers of words by males talking to males was associated with reduced task performance of male partners, it is arguable that males were less efficient communicators with other males in particular. Males were also less efficient communicators with female partners than female talkers because males used more words overall while achieving similar performance levels.

With respect to communicative efficiency, this analysis of number of words indicates that females were more efficient than males as talkers, using fewer words to evoke similar task performance levels from their partners, and that when a female talker used more words, her partner's performance level benefitted. Moreover, performance of females in relation to increased word usage by their partners was more consistently positive than that of males. Finally, performance of males did not benefit from increased use of words by male talkers, and their performance with

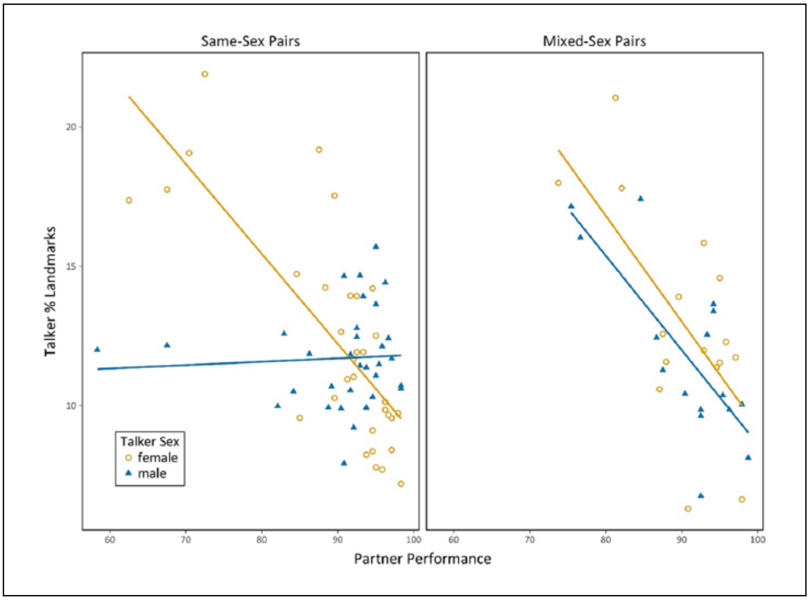

Figure 3. Correlations between a talker's percentage landmark words and their partner's task performance. The left panel displays data from same-sex pairs, and the right panel displays data from mixed-sex pairs. Female talkers are plotted with open orange circles and male talkers are plotted with filled blue triangles. Lines represent linear fits. The relationship for males in same-sex pairs differed from other pair types.

increased word usage by female talkers showed a weaker relationship. The next set of analyses probes these findings further, by examining the relationship between usage of landmark labels and task performance.

3.3.2 Landmark utterances and map task performance. An obvious prediction is that increased use of landmark label phrases in particular should be associated with enhanced performance on the map-matching task. However, results of analyses of landmark label phrase usage are counter-intuitive from this perspective, as shown in Table 2. Although correlations involving number of landmark words (which were included in counts of total numbers of words presented above) generally followed those of total numbers of words, the relationships were mostly weaker (except for males talking to females). In contrast, correlations involving percentage of landmark words relative to total words were strikingly different. As shown in Table 2, all significant correlations were negative, that is, an increased percentage of landmark words was related to lower task performance of partners, except in same-sex male pairs. A likely interpretation for this counter-intuitive finding is that relative increases in usage of landmark label phrases occur with pairs that are facing difficulties with the task. Another possibility is that production of relatively more non-landmark words includes those that would enhance precision in landmark placement, such as descriptions of locations. Figure 3 displays scatter plots for the relationship between the percentage landmark words produced by a talker and their partner's performance on the map task across same- and mixed-sex pairs of talkers.

Linear mixed effects modeling of these data assessed these correlation patterns and compared the relative strength of the associations between talker number of words, talker percentage landmarks, and partner task performance. In order to increase power in the dataset, these models used data that was split by maps within each pair (yielding 576 observations; the correlation analyses were based on 96 observations averaged by talker; performance increased modestly and evenly across maps within pairs). As before, continuous predictors were *z*-scale normalized and pair sex was treatment coded with male as the base level (vs. same and mixed). Modeling analyses revealed a three-way interaction between talker number of words, talker percentage landmark words, and pair sex as predictors of partner task performance (landmark number of words were not modeled because these data were correlated with number of words and showed similar but weaker effects). The only significant main effect in the fully interactive model was talker number of words, reflecting an overall tendency for greater numbers of words in male pairs to be associated with decreased partner task performance; ß = −2.64 (0.92), $t = -2.88$, $p = 0.004$. In contrast, the influence of percentage landmark words, which was significant when tested alone, ß = −1.84 (0.47), $t = -3.93$, $p < 0.0001$, was not a significant predictor when taking into account the influence of talker number of words, pair sex, and their interactions (likewise for pair sex). A three-way interaction with pair sex confirmed the observation that male pairs differed from other pairings with respect to the relationship between talker number of words, talker percentage landmarks, and partner task performance; $ß_{female}$ = 4.38 (1.47), $t = -2.97$, $p = 0.003$, $ß_{mixed}$ = 3.18 (1.49), $t = 2.14$, $p = 0.03$. Finally, the strongest predictor was the interaction between total number of words and female pair type (displayed in Figure 2), again reflecting the distinction of male pairs from female pairs; $ß_{female}$ = 5.14 (1.48), $t = 3.47$, $p = 0.0006$. Model comparisons confirmed that the three-way interactive model was a better fit to the data than additive models, models that excluded one of the predictors, and a base model with no predictors.

These analyses reveal that average number of words and percentage landmark words produced by talkers was associated with task performance of their partners in opposite ways, with greater numbers of words associated with increased performance levels of most partners, and greater percentages of landmark words associated with decreased performance. The loss of significance for percentage landmark words in the full model indicates that number of words was a stronger predictor than percentage landmark words. Interactions with pair sex confirmed that the influence of these factors in same-sex male pairs opposed those in other pair types. That is, same-sex male pairs showed a weak negative association between talker number of words and partner task performance, and performance of male partners in general was not positively related to variation in talker numbers of words and landmark utterances. These findings indicate that male talkers were not communicating as efficiently as female talkers because they averaged more words than females while their partners achieved roughly equivalent average performance levels overall, and because pair-wise increases in the number of words male talkers used with male partners were related to reduced partner task performance. Likewise, performance of males on the map task showed weaker relationships with numbers of words uttered by their partners, whether male or female.

4 Conclusions and future directions

The Montclair Map Task (MMT) was designed to elicit balanced contributions from talkers in a naturalistic task-oriented conversational setting, and the corpus comprises balanced representation of male and female talkers in same- and mixed-sex pairings. These features enable the corpus to contribute to a more comprehensive understanding of conversational interaction in general, and of the role of talker and pair sex in particular. With respect to talkativeness, the pattern of sex differences observed in the MMT corpus align with other studies that find that males interacting with strangers are more verbose than females, in both same- and mixed-sex pairings. An additional contribution is the inclusion of a fine-grain measure of communicative efficacy that reflects performance of both members of an interacting pair, and that can be related directly to aspects of communicative style. This measure proved fruitful in revealing subtle differences between pair types with respect to efficiency of communication. That is, same-sex male pairs of talkers were less efficient than female and mixed-sex pairs from the perspective of group performance because they used more words overall to accomplish equivalent task performance, and across same-sex male pairs, use of more words by talkers was weakly associated with lower performance from partners. Likewise, associations between talker percentage landmarks and partner task performance for same-sex male pairs differed from female and mixed-sex pairs. Overall, it appears that aspects of communicative style in male pairs of talkers were associated with less efficient communication in this setting. These findings warrant further investigation into more subtle aspects of language use in this corpus that might differ in same-sex male versus female and mixed-sex pairings. For example, it would be useful to classify each utterance according to its functional relevance for task completion in order to assess the quality of communication across pairs more directly.

This corpus can inform a robust body of research on task-oriented referential communication (reviewed in Brennan, Galati, & Kuhlen, 2010; Krauss, 1987; Pardo, 2017; Schober & Brennan, 2003). For example, many previous studies investigating the development of referential terms for ambiguous figures in role-differentiated settings have assessed efficiency in communication, finding that talkers shorten and tailor referents for specific addressees (e.g., Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Krauss & Bricker, 1967; Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992). In the current corpus, efficiency in communication differed for same-sex male pairs, possibly reflecting a more general tendency for males to talk more than females. A more likely alternative is that males were communicating differently, and the corpus transcriptions can provide more detailed information regarding communication style to resolve this issue. Finally, the fine-grained segmentation of structural attributes in the MMT corpus transcriptions, including turns, inter-turn intervals, pauses, fillers, overlaps, and backchannels can support broader investigations into the dynamics of conversational interaction.

Acknowledgements

The authors are grateful for the assistance of Emily Pardo and Melanie Ward with map-matching accuracy assessment and for the assistance of Michael Apfelbaum, Tainah Torres, and Sara Parker with data collection. The authors also appreciate thoughtful comments provided by the action editor, Dr Michael Vitevitch, Dr Mark Pitt, and an anonymous reviewer. The corpus recordings, segmentation and transcription annotation files, map task materials, and participant demographic files are available at the Montclair State University Digital Commons Data Repository (http://digitalcommons.montclair.edu/mmt_corpus). Access to the corpus recording files can be arranged by contacting the first author, Dr Jennifer S. Pardo at pardoj@montclair.edu.

Funding

This work was supported by the National Science Foundation (grant number BCS-1229033).

ORCID iD

Adelya Urmanche https://orcid.org/0000-0002-4311-2390

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. Retrieved from https://catalog.ldc.upenn.edu/LDC93S12

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. New York, NY: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory & Language*, *68*, 255–278. DOI: 10.1016/j.jml.2012.11.001

Bates, D. M., Mächler, M., Bolker, B., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bates, D. M., Mächler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Dai, B. (2016). lme4: Linear mixed-effects models using Eigen and S4 classes. R package version 1.1–12. Retrieved from http://cran.r-project.org/web/packages/lme4//index.html

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

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 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, MA: Academic Press. DOI: 10.1016/S0079–7421(10)53008–1

Brennan, S. E., Schuhmann, K. S., & Batres, K. M. (2013). Entrainment on the move and in the lab: The walking around corpus. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1934–1939). Austin, TX: Cognitive Science Society.

Canavan, A., Graff, D., & Zipperlen, G. (1997). CALLHOME American English Speech LDC97S42. Philadelphia: Linguistic Data Consortium. Retrieved from https://catalog.ldc.upenn.edu/ldc97s42

Cashdan, E. (1998). Smiles, speech, and body posture: How women and men display sociometric status and power. *Journal of Nonverbal Behavior*, *22*, 209–228.

Cieri, C., Miller, D., & Walker, K. (2004). The Fisher Corpus: A resource for the next generations of speech-to-text. *Proceedings of the 4th International Conference on Language Resources and Evaluation*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.61.8327

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

Fisher, W., Doddington, G., & Goudie-Marshall, K. (1986). The DARPA speech recognition research database: Specifications and status. *Proceedings of the DARPA Workshop on Speech Recognition*. *Palo Alto, CA*. pp. 93–99.

Godfrey, J., & Holliman, E. (1993). Switchboard-1 Release 2 LDC97S62 Philadelphia: Linguistic Data Consortium. Retrieved from https://catalog.ldc.upenn.edu/ldc97s62

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). Tests in Linear Mixed Effects Models. R package version 2.0–33.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English: phonetics, Phonology, and sound change*. Berlin, Germany: Mouton de Gruyter.

Labov, W., Rosenfelder, I., & Fruehwald, J. (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, *89*, 30–65. DOI: 10.1353/lan.2013.0015

Leaper, C., & Ayres, M. M. (2007). A meta-analytic review of gender variations in adults' language use: Talkativeness, affiliative speech, and assertive speech. *Personality & Social Psychology Review*, *11*, 328–363. DOI: 10.1177/1088868307302221

McMillan, J. R., Clifton, A. K., McGrath, D., & Gale, W. S. (1977). Women's language: Uncertainty or interpersonal sensitivity and emotionality? *Sex Roles*, *3*, 545–559.

Mulac, A. (1989). Men's and women's talk in same-gender and mixed-gender dyads: Power or polemic? *Journal of Language and Social Psychology*, *8*, 249–270.

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

Pardo, J. S. (2017). Parity and disparity in conversational interaction. In E. Fernández & H. S. Cairns (Eds.), *Handbook of psycholinguistics* (pp. 131–152). New York, NY: Wiley-Blackwell.

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*, 2254–2264. DOI: 10.3758/APP.72.8.2254

Pardo, J. S., Urmanche, A., Gash, H., Wiener, J., Mason, N., Wilman, S., …Decker, A. (2018). *The Montclair Map Task corpus of conversations in English*. Retrieved from http://digitalcommons.montclair.edu/mmt_corpus

Pardo, J. S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., & Ward, M. (submitted). A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics*.

Pitt, M., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication*, *45*, 89–95. DOI: 10.1016/j.specom.2004.09.001. Retrieved from http://buckeyecorpus.osu.edu

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random random effects and with binomial data. *Journal of Memory & Language*, *59*, 413–425.

R Development Core Team (2016). R: A language and environment for statistical computing, version 3.3.2. Retrieved from http://www.R-project.org/

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696–735.

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, NJ: Lawrence Erlbaum Associates.

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

Simkins-Bullock, J. A., & Wildman, B. G. (1991). An investigation into the relationships between gender and language. *Sex Roles*, *24*, 149–160.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native- and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language & Speech*, *53*, 510–540. Retrieved from http://groups.linguistics.northwestern.edu/speech_comm_group/wildcat/

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory & Language*, *31*, 183–194.

Wood, M. (1966). The influence of sex and knowledge of communication effectiveness on spontaneous speech. *Word*, *22*, 112–137.

Appendix 1: Demographic questionnaire

Montclair State University

Speech Communication Laboratory
Demographic questionnaire

1. Sex

What is your sex?
○ Male
○ Female

2. Age
How old are you? ____

3. Place of Origin & Residence(s)

Please list the places you have lived from birth until the current date, and indicate the number of years residing in each location.
City, State/Country Year Start Year End

4. Language Background

Do you have any training or exposure to a language other than English? If so, please list the language(s) below with the age at which you started learning/using the language(s) and your current fluency level.
(either fluent, basic conversational, or school setting only)
Language Year Start Fluency?

5. Ethnicity

Please specify your ethnicity.
○ Hispanic or Latino
○ Not Hispanic or Latino

6. Race

Please specify your race.
○ American Indian or Alaska Native
○ Asian
○ Black or African American
○ Native Hawaiian or Other Pacific Islander
○ White

7. Family Education Levels

What is the highest degree or level of school completed by either one of your PARENTS?
○ No schooling completed
○ Nursery school to 8th grade
○ 9th, 10th or 11th grade
○ 12th grade, no diploma
○ High school graduate—high school diploma or the equivalent (for example: GED)
○ Some college credit, but less than 1 year
○ 1 or more years of college, no degree
○ Associate degree (for example: AA, AS)
○ Bachelor's degree (for example: BA, AB, BS)
○ Master's degree (for example: MA, MS, MEng, MEd, MSW, MBA)
○ Professional degree (for example: MD, DDS, DVM, LLB, JD)
○ Doctorate degree (for example: PhD, EdD)

8. Family Household Income

What is your family's total household income per year?
○ Less than $10,000

- $10,000 to $19,999
- $20,000 to $29,999
- $30,000 to $39,999
- $40,000 to $49,999
- $50,000 to $59,999
- $60,000 to $69,999
- $70,000 to $79,999
- $80,000 to $89,999
- $90,000 to $99,999
- $100,000 to $149,999
- $150,000 or more

Appendix 2: Sample map task map pair

The images in Figure 4 comprise a single pair of maps used in the Montclair Map Task. Each map contains five shared landmarks and five unique landmarks. For example, both maps contain a *pyramid*, but Figure 4(a) on the left has a *telescope* at the upper left corner, while Figure 4(b) is missing this landmark. Instructions informed participants about the disparity in the composition of the landmarks, and asked them to discover the five missing landmarks that appear on their partner's map and draw a marker for each landmark in its corresponding location on their map. All talker pairs completed the map task with a total of six pairs of maps like those in Figure 4.

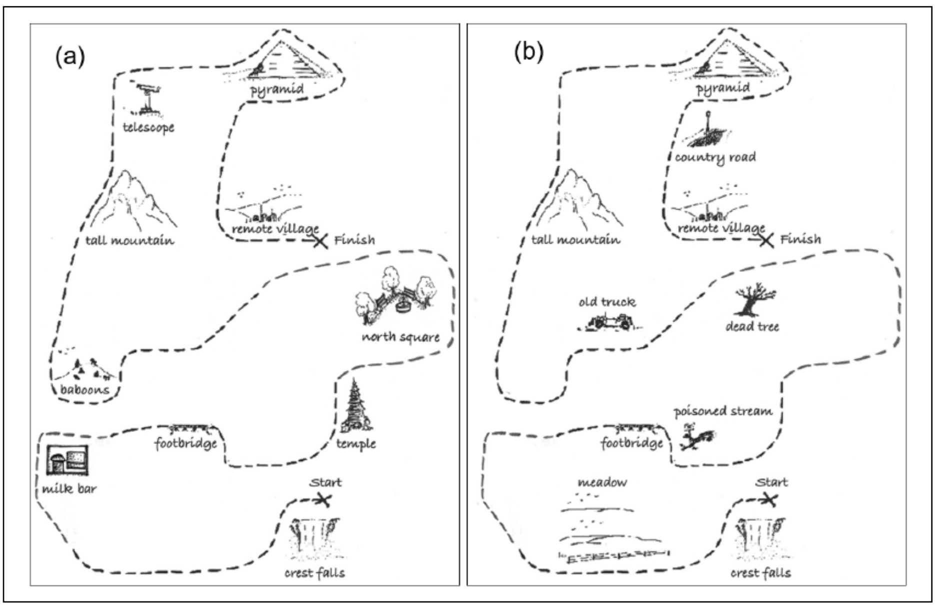

Figure 4. A single pair of maps used in the Montclair Map Task.

Appendix 3: Landmark label phrases

A listing of the full set of 79 landmark label phrases that appear across all six pairs of maps in the Montclair Map Task.

baboons
bakery
bear cave
blacksmith
buffalo
cactus
camera shop
caribbean palm
cattle ranch
cement roof house
chapel
cottages
country road
crest falls
dead tree
diamond mine
east lake
fallen rocks
farmed land
flat rocks
flowing river
footbridge
forked stream
fortress
garage
ghost town
golf course
graveyard
greasy wash water
green bay
headstone
huge nuclear plant
land parcels
large house
large cottage
lighthouse
marsh land
meadow
milk bar
monastery
monument
mud hut
museum
north square
oily rag
old truck
old mill
orange car
parked van
picket fence
pine forest
pirate ship
poisoned stream
pyramid
remote village
round hills
sandy shore
small forest
small island
stone cliffs
stony creek
tall mountain
tall pine
tavern
teepees
telephone booth
telescope
temple
totem pole
tower
trailer park
train crossing
train bridge
walled city
west lake
wheat field
winter garden
yacht club

Appendix 4: Correlation analyses for subset of data

Table 3 displays results of correlation analyses for a subset of data in which partner performance exceeded 80 points. In comparison with the data displayed in Table 2, the general pattern of correlation data is similar to that observed in the full dataset. In particular, the correlations for same-sex male pairs of talkers differ from same-sex female and mixed-sex pairs. However, the removal of nine data points for pairs with partners whose performance on the map task was below 80 points reduces power to the extent that many of the correlations are not significant in this subset. A new mixed-effects model analysis on the data subset revealed that the correlation between talker number of words and partner task performance differed between same-sex male and mixed-sex pairs ($ß_{mixed}$ = 1.59 (0.72), t = 2.22, p = 0.03), but no other main effects or interactions were significant. Most importantly, the three-way interaction revealed in the full dataset is no longer significant in the subset, ß = 1.03 (0.76), t = 1.363, p = 0.17.

Table 3. Correlations between talker words and task performance of partner; performance > 80 points.

	Male pairs	Female pairs	Mixed pairs	Males to females	Females to males
Correlation with task					
r number of words	*0.08*	0.43**	0.37**	*0.40*	*0.31*
r landmark words	*−0.02*	*0.17*	*0.22*	*0.33*	*0.06*
r percentage landmarks	*0.22*	−0.61**	−0.60**	−0.56**	−0.60**
Averages					
Partner task performance	92 (4.2)	93 (3.70)	92 (4.68)	91 (5.17)	93 (4.14)
Number of words	2,363 (879)	2,321 (810)	2,252 (1062)	2,592 (1230)	1,935 (793)
Number of landmark words	274 (103)	244 (52)	245 (75)	266 (78)	225 (68)
Percentage landmark words	11.7 (1.80)	11.3 (2096)	11.9 (3.31)	11.1 (2.64)	12.7 (3.77)

**$p < 0.05$.