5-2021

Gender Differences in Communication : A Task-Oriented Corpus Analysis

Bimash Budha
Montclair State University

Abstract

The current study explored the usage of various categories of words within the Montclair Map Task Corpus (MMTC) to determine if any relationships between the use of a certain category of words and task performance existed. The MMTC utilized a collaborative map task to explore the relationship between various patterns of communication styles and task performance. The current study built upon those goals by engaging in an in-depth analysis of the category of words utilized within the corpus. Based on the prior research on gender differences, it was hypothesized that category word usage would vary across different pair sex (male, female, mixed), and the relationship between their usage and partner performance would also vary. Based on the demand characteristics of the map task, certain categories such as direction and figure were expected to predict partner performance. No significant variations within the category usage across different pair sex were found, although the category types did vary in their proportion of use, overall. Affirm, direction, and figure category usage significantly predicted a partner's performance on the task. Overall, female pairs and mixed-sex pairs differed from male pairs such that the relationship between affirm, direction, and figure category usage was weaker in male pairs compared to female pairs and mixed-sex pairs.

MONTCLAIR STATE UNIVERSITY

Gender Differences in Communication: A Task-Oriented Corpus Analysis

by

Bimash Budha

A Master's Thesis Submitted to the Faculty of

Montclair State University

In Partial Fulfillment of the Requirements

For the Degree of

Master of Arts

May 2021

College of Humanities and Social Sciences

Department of Psychology

Thesis Committee:

[redacted]
Thesis Sponsor: Dr. Jennifer Pardo

[redacted]
Committee member: Dr. Alan Pehrson

[redacted]
Committee Member: Dr. John Paul Wilson

GENDER DIFFERENCES IN COMMUNICATION: A TASK-ORIENTED CORPUS ANALYSIS

A THESIS

Submitted in partial fulfillment of the requirements

For the degree of Master of Arts

by

Bimash Budha

Montclair State University

Montclair, NJ

2021

Table of Contents

Gender Differences in Communication: A Task-oriented Corpus Analysis

Language use primarily occurs in the form of speech conversations, and they involve an interaction between speakers and listeners. As a central aspect of our civilization, language use has captured the curiosity of philosophers, linguists, and psychologists for a long time. Our capacity to exchange ideas, information, and maintain social harmony through language is what draws attention, to be more precise. Therefore, researchers have studied the use of language extensively. Previously, communication studies have included corpora of naturally occurring speech, speech in controlled settings, and analyzed written texts as well. Although studying naturally occurring speech would be the most appropriate way to understand language and how it is used, the vastness and the randomness associated with such speech make it almost impossible to study language well. Analyses of written texts eliminate the extraneous variables associated with the speech in a natural setting however, written texts lack spontaneity and conversational component. Therefore, speech within a controlled setting is usually utilized in order to obtain functional and more meaningful results out of communication studies.

Such a wide array of corpora of speech and language use allow researchers to analyze various stylistic patterns in speech. Some communication studies have identified some specific nuances of speech such as shifts in the accent of partners upon interaction, explained by a theory popularized in the field as the communication accommodation theory (Giles et al., 1973). According to the speech accommodation theory proposed by Giles et al. (1973), an individual's speech is affected by the interaction with a conversational partner, the interaction's topic, and setting. Furthermore, speech accommodation theory has led communication research into a wide foray of research areas such as phonetic convergence research (Giles et al., 1991; Pardo et al., 2018). Speech communication research, therefore, has a plethora of studies on the effect of

partner, topic of the conversation, and setting in which it takes place on the speech behavior of an individual. Particularly, the interest in the effect of partners on communication has also taken the form of studies of gender differences in communication. For instance, statistical models predicted a social media user's gender with accuracies exceeding 90% based on their use of language, implying that there are gender differences (Schwartz et al., 2013).

Gender Differences in Language Use

Men and women seem to communicate differently, as many communication studies have focused on finding such differences in communication styles and communication goals (Leaper & Ayres, 2007). For instance, Mulac et al. (2001) investigated the gender-as-culture, or "two cultures" hypothesis proposed by Maltz and Borker (1982), which explains male/female differences in language use. They identified 16 language features that consistently predicted gender and tested them within four dimensions of intercultural style proposed by Gudykunst and Ting-Toomey in 1998 (as cited in Mulac et al., 2001). Language features such as directives, locatives, judgmental activities, references to quantity, etc. were identified as predictive of male language whereas references to emotions, questions, intensive adverbs, etc. were identified as female language features. These features were further tested within the four dimensions: direct versus indirect, succinct versus elaborate, personal versus contextual, and instrumental versus affective. Mulac et al.'s (2001) results indicated that directives, which was identified as a male language feature, were rated more direct than females' use of verbs. They concluded that male language features were rated as more direct, succinct, personal, and instrumental whereas female language features were rated as more indirect, affective, and elaborate.

Contrary to popular belief, research has also found that men were more talkative than women i.e. men speak more words, use more speaking turns, and sometimes speak at a faster rate

(Leaper & Ayres, 2007; Pardo et al., 2019). Hall (1984) and James and Drakich (1993) found that this tendency of men appeared more in mixed-gender pairs in task-oriented settings whereas such gender difference was unlikely in non-task-oriented contexts (as cited in Leaper & Ayres, 2007). If the gender differences do exist in a manner such that men speak more than women, and they also speak in an instrumental manner, it would imply that men would perform better in a task-oriented setting while pursuing specific instrumental goals.

Leaper and Ayres (2007) also distinguished between affiliative and assertive speech in an attempt to study gender differences in language use. Affiliative speech includes expressing agreement whereas assertive speech includes providing information amongst many other examples. Leaper and Ayres (2007) found that men used more assertive speech and women used more affiliative speech. Socialization theory and social constructionist/contextualist approach explain such gender differences (Leaper & Ayres, 2007)

The socialization explanation focuses on the cumulative influences of gender-segregated activities as a child on an adult's communicative style (Maltz & Borker, 1982). Based on this theory, women utilize more affiliative speech since they engage in cooperative social play with their peers while men utilize more assertive speech because they participate more in isolative play or competitive play with their peers. Maltz and Borker (1982) proposed that girls utilize their words such that it creates and maintains closeness with peers and others in a supportive and inclusive manner whereas boys utilize commands, challenging statements, and questions to assert their position of dominance.

On the other hand, the social constructionist (or contextualist) approach explains the gender differences in language use as a function of demand characteristics of the situation therefore, interactive context is vital to explaining gender differences in this paradigm. For

instance, gender differences in talkativeness were more likely within task-oriented contexts than non-task-oriented contexts (Hall, 1984, James & Drakich, 1993, as cited in Leaper & Ayres, 2007). Furthermore, the social context in most cases demands higher status and power for men and usually results in a subordinate status of women which is conveyed through their use of language (Hall, 2006). Carli (1990), thus, suggested that the gender difference would be most likely in a mixed-gender composition since the status disparity is most stark and salient (as cited in Leaper & Ayres, 2007).

Consistent with these theories, gender differences may also exist in the manner men and women use actual words. For instance, men's more succinct and instrumental manner of language use would lead them to use object-oriented, instrumental, and task-related words more. On the other hand, women's subordinated role within a social context would lead them to use subtle, passive, and non-threatening words.

Word Categories in Language use

Beginning with Lakoff's (1975) investigation of gender differences at the level of phrases, gender communication research has considered linguistic categories. One of the key findings from Lakoff (1975) was that females used specific phrases such as hedges (e.g., "it seems like") and tag questions (e.g., "…aren't you?"). Some studies have found greater use of such phrases amongst women (Newman et al., 2008). Similarly, uncertainty verb phrases such as "I wonder if" occurred more in women's speech compared to men (Hartman, 1976).

Since Lakoff's (1975) work, gender differences have also been examined by researchers studying actual words used by people. For instance, research suggests that women use more intensive adverbs, conjunctions, and auxiliary verbs whereas men use more articles and references to location (Biber et al., 1998, as cited in Newman et al., 2008). In addition, Schwartz

et al. (2013) found that females use more emotion words and first-person singular pronouns whereas males use more swear words and object references such as “Xbox”. Similarly, Newman et al. (2008) studied gender differences in language use by using standard linguistic categories. The authors utilized Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2001), which analyzes text on a word-by-word basis and classifies them into one of 74 linguistic categories after comparing them with a dictionary of over 2,000 words. They utilized a database of over 14,000 text files from 70 separate studies in this study. Thus, computerized linguistic investigations became a methodological development in communication research. They found that women used words related to psychological and social processes (such as “happy”, “joy”, “pride” etc.) whereas men used words related to object properties and impersonal topics. Therefore, phrase-level findings and word-level findings have further supported past findings from the study of language features by Mulac et al. (2001).

Similarly, Park et al. (2016) studied a large social media dataset of approximately 67,000 Facebook users. Across two studies, they investigated gender differences in language use by topics, and within interpersonal dimensions of assertiveness and affiliation. The researchers were examining the gender differences in the use of words/topics between males and females. Study 1 did not include any definitive hypothesis. Instead, the researchers were interested in identifying commonly used topics, therefore, the analysis was exploratory in nature. In order to achieve this, they drew participants from a third-party Facebook application known as MyPersonality. Participants provided their written consent to the anonymous use of their responses and access to their status updates. The selection criterion for the studies included the users who granted MyPersonality application access to their post status, wrote more than a thousand words, and

provided their gender and age, which was between 16 and 64 years. Study 2 included participants who, in addition, completed a 100-item personality measure.

In study 1, topic-based linguistic analyses of gender differences were carried using latent Dirichlet allocation (LDA) to identify topics. LDA is an unsupervised clustering algorithm that finds potentially meaningful clusters of words in large samples of natural language (Blei et al, 2003, as cited in Park et al., 2016). Automatic screening of topics led to 1281 unique LDA topics, which excluded semantically similar topics. The effect size of the occurrence of topics in either male or female participants was 0.12 on average. Despite the low effect sizes, the automatically identified topics by LDA were gender-linked language topics. For instance, most female-linked topics were the intensive adverbs such as "so" or "ridiculously" whereas male-linked topics were references to sports and occupations such as "management" or "business".

In study 2, the topics identified in study 1 were assigned a psychological label associated either with affiliative or assertiveness. The researchers achieved this by examining correlations between topic use and self-reported personality measures of extraversion (assertiveness) and agreeableness (affiliation). The reported measure was taken using a 100-item Big Five measure consisting of items from the International Personality Item Pool (IPIP). Age and gender were controlled to derive genuine personality differences in the topic use, and further controlling either extraversion or agreeableness ensured that the high degree of correlation between these was taken into consideration. Study 2 revealed that women used more affiliative language however, they also used slightly more assertive language whereas males used colder, more hostile, and impersonal language.

They also found that topics such as friends, family, and social life were most associated with females whereas swearing, anger, discussion of objects instead of people, and the use of

argumentative language was most associated with males. Also, females used highly affiliative language more, however, they also used highly assertive language. In other words, females used highly assertive language that was also interpersonally warmer whereas males' assertive language was colder. Park et al. (2016) utilized approaches from computational linguistics and psychological theory, and it allowed them to determine affiliative and assertive topics based on their association to user's extraversion and agreeableness scores. One limitation of their study was the use of social media samples, which is different from previous contexts used in language studies. The future implications from Park et al. (2016) include, but are not limited to, utilizing automatic labeling as a means for testing psychological theories i.e. open vs closed vocabulary approach. More importantly, their methods highlight a possibility for communication research of studying word category usage as an attribute of conversational style.

As discussed previously, naturally occurring speech conversations in task-oriented settings provide a more accurate understanding of the language use and the plethora of nuances associated with it. For instance, Koulouri et al. (2017) studied gender differences in communication in complex, collaborative tasks and how it influenced task performance. Koulouri et al. (2017) did not find any performance difference between males and females, however, differences were observed in their communication styles. Males tended to utilize novel vocabulary in the event of a miscommunication whereas females utilized a more familiar speech behavior. It is important to note that Koulouri et al. (2017) is among very few studies that have attempted to associate language features with performance in an objective task.

In an adaptation of a conversational map task designed to collect speech in a non-scripted setting, Pardo et al. (2018) collected speech of either males, females, or mixed pairs of both genders involved in a role-neutral map task. One of the objectives of the Montclair Map Task

Corpus (MMTC) project was to investigate the relationship between task performance and various attributes of conversational style (Pardo et al., 2018). Some findings denoted gender differences in the relationship between task performance and the various attributes of communication style such as the total number of words used, proportion of overlaps, and the number of turns taken. For instance, male talkers' use of more words was negatively correlated with their male partners' task performance whereas females' use of more words was positively correlated with their female partners' performance (Pardo et al., 2018). Also, it is equally important to note that within mixed-sex pairs, male listeners performed worse compared to female listeners when their partners used more words.

Past research suggests that men used language with the instrumental purpose of conveying relevant information, therefore, men should perform better on a task that requires the use of spatial features. For instance, Newman et al. (2008) reported that men use more spatial words such as "above" and "over". Also, across empirical studies and literature reviews by Mulac et al. (2001) and Mulac (2006) references to location and quantity were found to be male-linked features. Since the correlation between the total number of words produced by a talker and their partner's performance in the task was negative in same-sex male pairs, further analyses need to be carried out in order to understand Pardo et al.'s (2019) findings.

Montclair Map Task Corpus

Pardo et al. (2019) collected a speech corpus from conversations between pairs of same-sex and mixed-sex pairings. The Montclair Map Task (MMT) is a role-neutral version of the HCRC map task from Anderson et al. (1991). The original HCRC task required participants to collaborate in order to recreate a map route printed in the other participant's map (Anderson et al., 1991). It produced unscripted task-oriented language however, the nature of the task led to

uneven production of speech by participants. In other words, the participants' roles within the task created disparities within the speech collected. The Montclair Map Task, in contrast, required participants to collaborate in order to place missing landmarks for each participant in a pair across six maps (Pardo et al., 2018).

The more talkative member of each pair on average produced 30% more words in total when compared to 2.7 times more words produced by the givers in HCRC Map Task utilized by Pardo (2006), therefore, the MMTC achieved more balanced conversations when compared to other corpora (as cited in Pardo et al., 2019). A total of 96 participants were randomly assigned to either same-sex male, same-sex female, or mixed-sex pairing, resulting in a total of 48 pairs (16 male, 16 female, and 16 mixed pairs). Each pair was assigned six maps, and each member of a pair received the maps with five shared and five unique landmarks.

The transcribed speech from the conversations was analyzed to assess gender differences in the efficacy and efficiency of communication related to the performance scores for partners. The placement of landmark markers on the maps indicated relatively high accuracy among participants in locating missing landmarks. Therefore, the placement of landmarks was rated on a 100-points scale based on the precision with which the participants achieved the goal of landmark placements. In general, precision scores increased from Map 1 to Map 6. The variability in the precision scores was indicative of the efficacy of communication.

In addition, corpus events such as turns, words, syllables, landmarks, pauses, fillers, overlaps, and backchannels were counted and also split for talker sex and pair sex information. The analysis revealed that male talkers produced more instances of all corpus event types than female talkers, except that females paused slightly more often than males. Across pair sex types, male pairs produced more instances of corpus event types than female pairs, except female pairs

paused more, and that mixed-sex pairs produced more fillers and backchannels. More analyses revealed the efficiency of these stylistic differences in communication.

Efficiency was determined as the function of communication style on the partners' task performance, which was measured using the precision scores, or in other words, their partner's efficacy. Analysis revealed that males produced more words overall, and produced speech at a faster rate. One interesting finding associated with the efficiency of their speech was that there was a trend towards partners performing worse when more words were produced, $r(30) = -0.33$, $p = 0.06$. In contrast, female pairs had a strong positive correlation between the number of words produced and partner map performance, $r(30) = 0.60$, $p = 0.0003$. Similarly, there was a positive correlation between the number of words produced and partner map performance in mixed-sex pairs, $r(30) = 0.48$, $p = 0.008$. In other words, males were found to be less efficient communicators than females with respect to partner map task performance. It also means that the male listeners performed worse compared to female listeners when their partners used more words. However, male listeners performed better when female partners produced more words compared to male partners. Also, the number of words spoken by same-sex male pairs reduced dramatically from map 1 to map 2 whereas their partner performance increased.

To fully understand these patterns observed within the Montclair Map Task Corpus, further detailed analyses of the corpus must be conducted. For instance, further differences in communication styles may be contributing towards the varying partner performance scores between the pairs and between the talkers. In order to understand the variability observed within the first two maps within the MMTC, Bell (2020) investigated each talker turns including a question, and the turns determined as questions were coded into further question types. First, the individual turns were coded into one of five functional categories: continuation, answer,

question, redirect, and attempt. Since question and answer turn types were a significant portion of all speaker turns, questions were further coded into categories based on speakers' goals such as challenge, clarification, information, repetition, direction, and comparison. Bell (2020) hypothesized that males would utilize challenge questions more than females, especially across pair sex, and females would utilize information seeking and clarification seeking questions. The hypothesis was based on research regarding affiliative and assertive speech in men and women. Pair sex variations in question types were expected however, no consistent gender effects were found (Bell, 2020). Since the turn-by-turn coding did not produce any meaningful results to explain gender differences in map task performance, further detailed analysis of the MMTC is necessary.

The Current Study

The current study builds upon the prior investigations into the Montclair Map Task Corpus. Pardo et al. (2019) provided the analysis of stylistic differences within the corpus, and measures of the relationship between various speech components and their effect on a partner's map task performance. For instance, male talkers produced more words, turns, and overlaps than female talkers in both same-sex and mixed-sex pairings. Pardo et al. (2019) also found a trend towards a negative relationship between the total number of words produced by a talker and their partner's performance in the map task in same-sex male pairs. In contrast, same-sex female pairs performed significantly better when a talker produced a greater number of words, and the correlation was similar in mixed-sex pairs, however, it was smaller than in same-sex female pairs. Upon further investigation into the talker's sex, Pardo et al. (2019) found that the correlation between male talkers' total number of words and their female partner's performance

was very strong whereas male partner's performance was not strongly correlated with a female talker's production of words.

Bell (2020) investigated these gender differences in affiliative and assertive language use by coding the types of questions utilized by the talkers within the first two maps. No significant relationships were found between the types of questions and partners' task performance. Bell's (2020) study did not include an analysis of the complete six sets of maps. The current study, therefore, will engage in a more in-depth analysis of the corpora speech utilizing language processing techniques such as bag-of-words analysis of events delineated as context-speech by Pardo et al. (2019). Specifically, the current study will investigate the use of topic words and their association with task performance, and if such association exists in a manner that explains the gender disparities observed in Pardo et al. (2019).

The current study explores gender differences, both between pairs and between individual speakers, found in the MMTC by focusing on the word categories utilized within the conversations. Prior coding of the corpus transcription had identified various corpus events such as landmarks, pauses, fillers, overlaps, backchannels, non-speech, and context. Any utterances produced by a partner during the main talker's turn that did not initiate a new turn e.g., uh-huh, okay, etc., were coded as backchannel events. Similarly, pause fillers such as "um" and "uh" were coded as filler events. Any other speech events within a turn that was not part of a landmark label phrase were coded as context. In other words, context speech includes any meaningful speech produced within the task. Pardo et al. (2019) had found a trend towards a significant negative relationship between the total number of words produced by a talker and their partner's performance in the map task in same-sex male pairs whereas same-sex female

pairs had a significant positive relationship. Therefore, an in-depth analysis of the meaning of the words used within the context speech might explain Pardo et al.'s (2019) findings.

Relevant to the significance of the meaning of words on partner performance, only context speech was further coded into individual word categories. Since prior analyses did not find any significant differences in the use of non-context speech corpus events such as backchannels and landmark label phrases, the current study will analyze only the context speech further. Unlike Bell (2020), this study is aimed at understanding the use of various word categories in the task-oriented setting of the Montclair Map Task, therefore, all six of the trials were evaluated.

Words from context speech were categorized into one of thirteen different categories: affirm, verb, direction, figure, directive, landmark, pronoun, function, number, incomplete, compare, size, and other. Speech that conveys instrumental information necessary for the map task should assist in task completion. Based on the navigational requirement of the task, it is expected that the use of direction and figure terms will be correlated to the partner's performance on the map task. In other words, it is hypothesized that the relationships between a talker's use of categories and their partner's performance on the task will vary across pair sex. To test this, each word occurring in the context speech within the MMTC was coded into the relevant word categories.

Methods

Montclair Map Task Corpus

The current study utilized original coding from a previously recorded speech corpus, the Montclair Map Task Corpus. The procedures and methods used in collecting the corpus as well as the original coding of the corpus transcript is described below. In addition, the current study

utilizes novel coding, which is explained as well. For more information on the Montclair Map Task Corpus, see Pardo et al. (2019).

Participants

A total of 96 participants (48 males, 48 females), who were native English speakers with normal hearing and speech, were recruited from Montclair State University between March 2013 and May 2014. They completed IRB-approved informed consent along with a demographic questionnaire, for which they received US$20 as compensation. The participants were randomly assigned into gender pairs – 16 same-sex male, 16 same-sex female, and 16 mixed-sex. The recordings took place during March-May of 2013 for the 24 pairs whereas the other 24 pairs were recorded in May 2014.

The demographic questionnaire collected data about the participants' residential history, ethnicity, race, family income level, parents' education level, language background, age, and sex. Participants' age ranged from 18 to 38 years old, with a mean of 21 years old (SD = 2.8 years). Nearly all participants had resided in New Jersey for at least three years whereas nearly 90% of participants had been born in and lived most of their lives in New Jersey/New York/Pennsylvania area. Seven participants had been born outside of the United States, however, they had lived in New Jersey for at least five years at the time of recording. 20 of the participants identified themselves as Hispanic/Latino or Not Hispanic/Latino, and two participants left the question blank. Similarly, 61 participants identified themselves as White, 15 identified as Black, four as Asian, two as American Indian or Alaskan Native, one as Black/White, one as Native Hawaiian/White, one as Other and one wrote Latina (eight participants left the question related to participants' race).

Pertaining to the variable of interest to the current study, language backgrounds were collected with a final question. The participants had to list any languages they knew in addition to English, to indicate the year they began learning the language and their proficiency in that language (fluent, basic conversational, or school setting only). Based on the responses, 26 participants reported that they were fluent in a second language, and 19 participants left the item unanswered. Many of the individuals who listed a second language and indicated that they were fluent also indicated the year of their birth as the year they began learning the second language, thus, it is likely that those individuals would be considered bilingual.

Procedures

Participants were randomly paired into 16 same-sex male pairs, 16 same-sex female pairs, and 16 mixed-sex pairs. The pairs included talkers and partners who were not acquainted prior to the recording session, and they could not see each other during the conversational task since they were placed into two booths separated by a room divider. The booths were soundproof, and the participants communicated with each other through AKG head-mounted microphones connected to computers situated outside the booth. Participants used a pencil to complete the map task, and the conversations between the partners during the task were recorded using SoundStudio software onto an individual time-aligned channel in a two-channel audio file.

Conversational task

The HCRC Map Task (Anderson et al., 1991) was adapted into a modified version for the Montclair Map Task (MMT). The MMT involved paired iconic maps with labeled landmarks, and the participants must reconcile the differences between the paired maps. Each gender pair were assigned six paired maps. These paired maps consisted of five landmarks that

were shared in both and five landmarks that were unique to each map, see Appendix A. A total of 79 unique landmarks, with some landmarks appearing on multiple maps, were utilized.

The MMT required participants to find the missing landmarks and provide a location within their maps by drawing labeled markers as accurately as possible. Thus, the MMT is essentially a map-matching task. The goal of this map-matching task is to evoke natural conversational interaction and collect naturally occurring speech data generated by the participants, see Appendix B for a sample of their conversation. Since they were unable to see their partner's maps, this task ensured that they converse in order to complete the task. And, their conversation is a key component in the success of their task completion.

Map Task Accuracy

Two independent raters scored accuracy in locating missing landmarks. The precision of landmark placement was assigned 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 regardless of a path), and missing. A correct location was allocated 20 points, 15 points for adjacent, 10 points for nearby, 5 points for distant location, and 0 points for missing landmarks. The interrater reliability for this measure was high ($r = 0.84$), therefore, the raters' scores were averaged to comprise final precision scores.

Transcription Protocol

The transcriptions of all 48 conversations were carried out by trained research assistants, who also segmented the two-channel audio files. The annotation procedures in Praat Software were used for transcriptions. The audio files comprised of two time-aligned channels, one for each talker. The research assistants segmented the recording into delineations for map number (1-6) to indicate which map the particular speech was being produced on, speaking turns

(utterance by a single talker accomplishing a communicative goal; Sacks, Schegloff, & Jefferson, 1974, as cited in Pardo et al., 2019) to indicate which talker was speaking, a backchannel (utterances produced by a speaker during the main talker's turn that does not initiate a new turn, e.g., "uh-huh", "okay", etc.), overlapping speech (both talkers speak simultaneously), pauses (silent interval during a speaker's turn that is greater than 100 ms), fillers (fillers produced during pauses such as "uh", "um", etc.), and finally, context speech (words produced within a turn that were not part of a landmark label phrase). Praat textgrid files stored the data from segmentation, transcription, and labeling, and conversational analyses used data extracted from these files.

Word Categories Coding Protocol

The MMTC results from Pardo et al. (2019) indicated gender differences within multiple aspects of communication, and the results from Bell (2020) suggested a more focused approach to identify gender differences and their impact on partners' scores. In order to associate the task-oriented nature of the conversations and their relative impact on the success of the task, the MMTC was further coded. Since a majority of the task-relevant speech was uttered within the context speech segment, further coding was carried out on context speech.

First, trained research assistants filtered the transcription files for any context speech events. Thus, they deleted any other corpus event such as pauses, fillers, non-speech, backchannels, etc. for all 48 files (each file including conversations for each pair), while carefully saving the context speech events. Then, dfm function from the quanteda package in R statistical platform was utilized in order to generate a list of all the unique words used within the entire corpus for context speech events.

Second, the trained research assistants coded the unique words into one of 13 categories: affirm, verb, direction, figure, directive, landmark, pronoun, function, number, incomplete,

compare, size, and other. Based on the task-oriented nature of the context speech event, the words were categorized accordingly to reflect the map-matching task performance. Contemporary linguistic categories and task-relevant categories of words were generated based on past research. For instance, men utilized directives, locatives, references to quantity, and articles whereas women utilized more intensive adverbs, references to emotions, first-person singular pronouns, etc (Mulac et al., 2001; Newman et al., 2008; Schwartz et al., 2013; Park et al., 2016).

The words were coded as a *affirm* (words indicating approval such as "yup", "Yeah", "okay", "alright", "mhm", etc.), *verb* (words relating to any indication of an action unrelated to direction such as "see", "called", etc.), *direction* (words relating to any indication of direction such as "next", "inside", "side", "ahead", "bent", "straight", etc.), *figure* (words indicating any shapes and figures such as "line", "square", "dotted", "space", "angle", "curve", etc.), *directive* (words indicating any sort of assertive action or intent such as "do", "don't", "wrong", "no", etc.), *landmark* (words indicating partial landmark labels), *pronoun* (any pronouns such as first-person pronouns, "it", "that", "something", etc.), *function* (words indicating functionality in speech such as articles, prepositions, conjunctions), *number*, *incomplete* words, *size* (words indicating any sense of size information such as "little", "bit", "much", "big", etc.), and *other* (words indicating no relation to the map-task based language usage such as adverbs, "literally", "actually", "probably", etc). However, in an attempt to correctly code ambiguous words such as "right", which could either be an affirmative word, or a word indicating the direction to the right or any other category, were coded into a separate category *ambi*. The research assistants looked for each instance of such *ambi* words and further categorized them into their respective word categories.

Finally, the unique words and their respective categories were combined to the original context speech data file so that each file contained the word and its category information, and other relevant information such as which speaker uttered the word, respective map number, talker's sex, and pair sex.

Data Analysis

Data were manipulated and analyzed using the R statistical platform (R version 4.0.5 R Development Core Team, 2016). All analyses were set to an alpha level of 0.05, and they were exploratory in nature. To investigate any patterns in the context speech events within the MMTC, instances of word utterances were compared based on pair sex and talker sex.

First, the proportion for each category use was calculated by dividing the number of words in each word category type by the total number of words in each map for each talker, regardless of type. Since landmarks, incomplete, and other word categories were irrelevant to the purposes of data analyses, they were removed from any further data manipulation. To assess the effects of pair sex on category usage, a two-way repeated-measures ANOVA included the proportions of coding type (word categories) as a within-subjects factor and pair sex (male, female, or mixed) as a between-subjects factor.

In order to predict a partner's performance from the proportion of word usage, regression analyses were performed using lmer function from the lme4 package with t-test results obtained by using the lmertest package in R. The first linear mixed-effects regression analysis was performed to investigate if any category would predict a partner's map performance, with talker as the random intercept. Additional regressions were performed to isolate the significant predictors of the partner's map task performance. Three regressions were performed to predict partner's performance with the use of either affirm, direction, or figure category across pair sex.

Also, a regression was carried out to observe any effects of talker sex on the relationship between partner's performance and use of figure words.

Results

Overall, within context speech, male pairs produced a total of 68,403 words whereas females produced 61,271 words and mixed-sex pairs uttered 60,305 words. Similarly, in a downward trend, the total number of words across pair sex types was found to be 35,864 at map 1 and 31,874 at map 6. Similar to the findings of Pardo et al. (2019) on the total number of words, male talkers produced more words within context speech events than female talkers; males produced 102,177 while females produced 87,802 words. In order to analyze the data in more depth, further analyses were performed.

In order to test the hypothesis that the relationship between a talker's use of different word categories and their partner's performance on the map task varied across pair sex, the category proportions were calculated. For ease of presentation, the word categories have been split into high proportion word categories and low proportion word categories in Figure 1. Affirm, direction, function, pronoun, and verb word categories were classified as high proportion word categories while compare, directive, figure, number, and size were considered low proportion word categories.

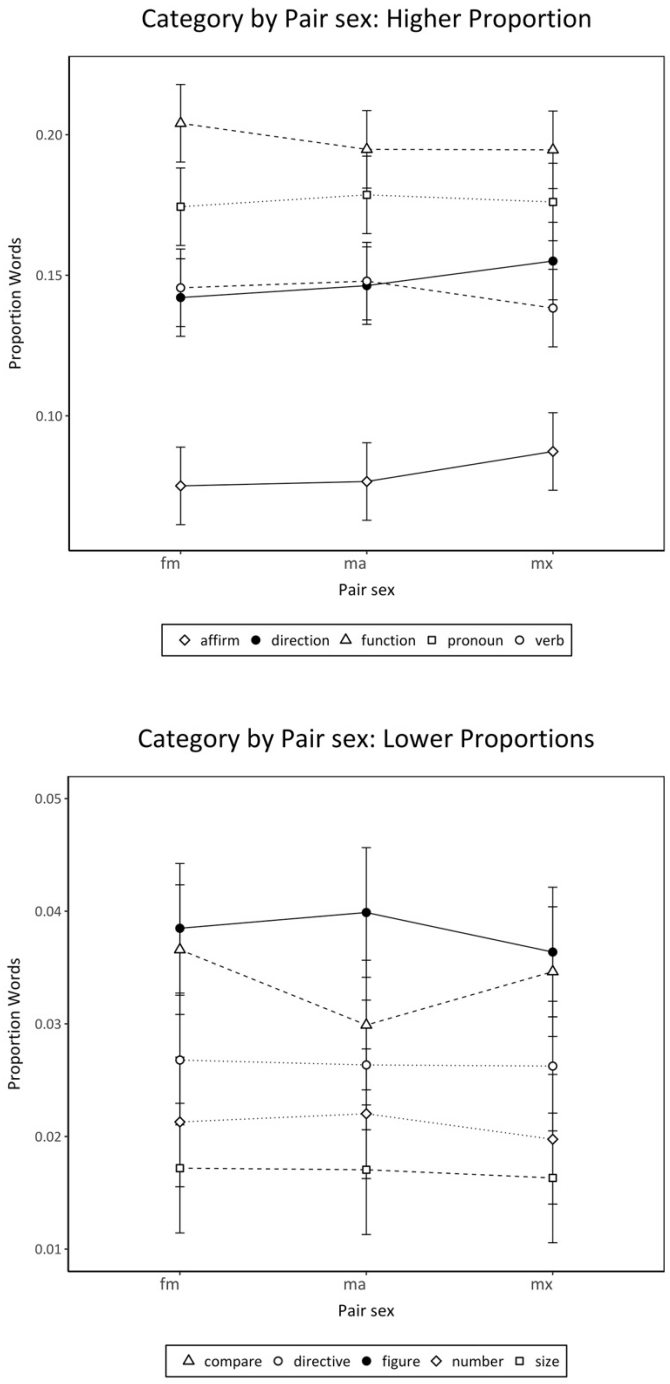

Figure 1. Proportion of each category type based on pair sex (female, male, and mixed-sex pairs). The plots have been separated by higher proportion category types (top) and lower proportion categories (bottom). Error bars represent a 95% confidence interval using Fischer's Least Significant Difference (FLSD).

There were differences in the use of the word categories themselves, i.e. some word categories were used in higher proportion than others, as seen in Figure 1. For instance, *affirm* word category occurred in a higher proportion than *compare* category across all pair sex types, which was a lower proportion category. Also, talkers used *function* category the most across all pair sex whereas *size* category was utilized the least. Within higher proportion categories, *direction* was used relatively less whereas *affirm* category seemed to be an intermediate between lower proportion and higher proportion categories. On the other hand, *figure* terms were towards the top of the lower proportion usage.

In order to confirm any differences observed within Figure 1, a two-way repeated-measures ANOVA was carried out to investigate any interaction between pair sex as a between-subjects factor and word category as a within-subjects factor. The two-way ANOVA assessed the influence of pair sex (male, female, mixed) and category (affirm, direction, function, pronoun, verb, compare, directive, figure, number, and size) on the proportion of word use. Mauchly's test for sphericity violation was significant, therefore, Greenhouse-Geisser adjustments were used instead. The results indicated a significant main effect of category of word usage [$F(4.03, 374.55) = 1079.47$, $p < 0.05$, $\eta^2 = 0.92$]. Conversely, the two-way interaction effect between category and pair sex was found to be non-significant [$F(8.05, 374.55) = 1.29$, $p = 0.247$, $\eta^2 = 0.03$]. Similarly, the main effect of pair sex on proportion of word usage was also non-significant [$F(2, 93) = 0.55$, $p = 0.58$, $\eta^2 = 0.0001$]. The complete results from the two-way ANOVA are available in Appendix C Table 1.

In other words, categories were significantly different from one another in their proportion of usage. However, the interaction between category and pair sex was not significant

which indicates that the previous findings that suggested category use would vary across pair sex did not replicate. Since the ANOVA results indicated variation in category types, the relationship between a talker's use of category types and their partner's performance on the map task needs to be further evaluated.

Word Category and Task Performance

The use of certain category types may be correlated with partner performance on map task. For instance, a greater use of *direction* words may assist in the map task, thus, it may be correlated to higher map task performance scores. In order to further investigate the observed differences in proportion use, the relationship between proportion usage and partner's map task performance must be analyzed. This investigation builds upon the previous findings that overall word usage was related to partner performance such that the relationship between male talker's use of a greater number of overall words was negatively related to their partner's performance, whereas the relationship was positive for female talkers. Furthermore, the current investigation examined whether differences in specific types of words can explain the finding. A linear mixed-effect regression model assessed which, if any, categories predicted a partner's map performance. Follow-up analyses examined any differences within pair types in the relationships between category proportion usage and partner's map task performance, and differences within talker sex in the relationships between category use and partner performance were also analyzed.

In order to normalize the scores, the proportion usage for each category was converted into z-scores using scale function from the base R package. The full mixed-effects regression model included the normalized scores for all word categories as predictors, and talker was included in the regression models as the random intercept, which is also depicted by the model formula in Appendix C Table 2.

The results indicated that only the *affirm*, *direction*, and *figure* category usage showed significant relationships with the partner's map task performance. In other words, variation across talkers in the usage of *affirm*, *direction*, or *figure* words was related to variation in their partner's performance in the map task. Partner performance was negatively predicted by talker's use of *affirm* words [β = -1.15 (0.50), t(537.86) = -2.31, p = 0.021]. In contrast, partner's performance was positively predicted by talker's use of *direction* words [β = 1.32 (0.54), t(564.6) = 2.44, p = 0.015], and talker's use of *figure* words [β = 2.05 (0.47), t(565.48) = 4.36, p < 0.01]. All the other word categories were found to be non-significant in their relationship to the partner's map task performance (all non-significant ts < 1.54). The results from the full model are available in Appendix C Table 2.

In order to further investigate the variation of significant predictor categories i.e. *affirm*, *direction*, and *figure* across different pair types (males, females, mixed-sex pairs), additional linear mixed-effects regression models were run for individual predictor categories with random intercept set to pair sex such that they included interaction terms between pair sex and each predictor category.

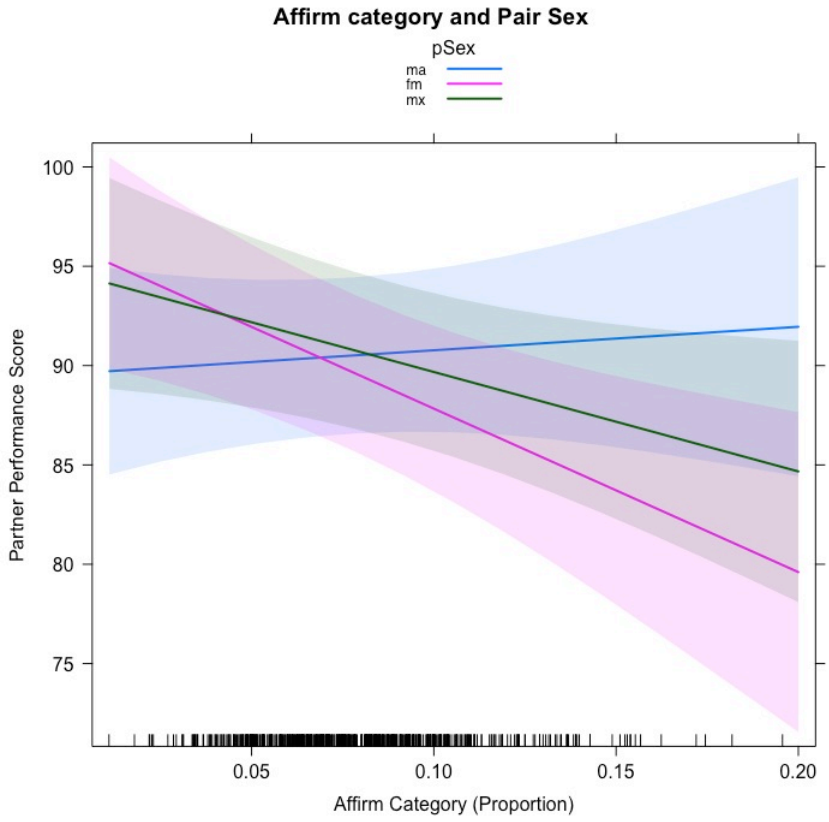

Figure 2. A plot of the variation across different pair types (male pairs, female pairs, mixed-sex pairs) in the relationship between proportion affirm category usage and partner map performance. Fixed effects from the linear-mixed effects regression model plot relationships for male pairs in blue, female pairs in magenta, and mixed pairs in green. The shaded regions represent a 95 % confidence interval.

As evident from Figure 2, the relationship between partner map performance and affirm category usage of a talker is negative in same-sex female pairs and mixed-sex pairs. In contrast, male pairs show no relationship between affirm category usage and partner performance in the map task. The mixed-effects regression analysis that included an interaction term between pair sex and affirm category predictor confirmed the observation that same-sex male pairs differed from same-sex female pairs [β = -2.59 (0.98), t(556.85) = -2.41, p = 0.016]. Conversely, same-

sex male pairs did not differ significantly from mixed-sex pairs [β = -1.70 (0.98), t(566.94) = -1.73, p = 0.084].

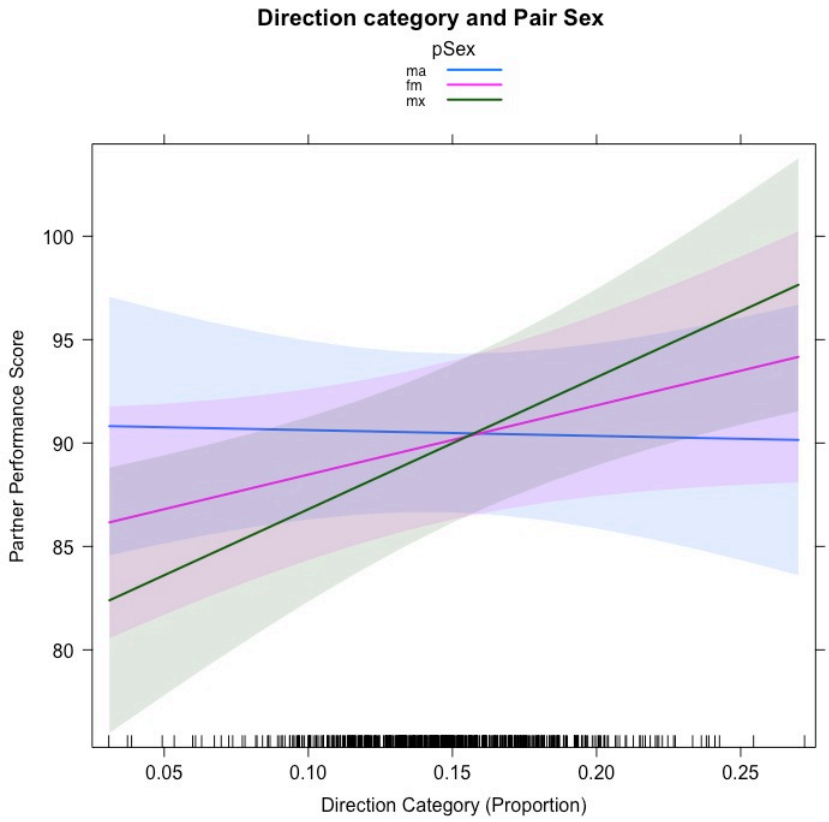

Figure 3. A plot of the variation across male pairs (ma), female pairs (fm), mixed-sex pairs (mx) in the relationship between direction category usage and partner map performance. Fixed effects from the linear mixed-effects regression model plot relationships for male pairs in blue, female pairs in magenta, and mixed pairs in green. The shaded regions represent a 95 % confidence interval.

As depicted in Figure 3, the relationship between partner map performance and direction category usage of a talker appears to be positive in same-sex female pairs and mixed-sex pairs. Again, same-sex male pairs show no relationship between partner performance and direction terms usage. In addition, the slopes indicate that same-sex male pairs differed slightly from

same-sex female pairs whereas the difference was bigger between same-sex male pairs and mixed-sex pairs. The mixed-effects regression that included an interaction term between direction category use and pair sex confirmed that the male pairs differed from the mixed-sex pairs [$\beta = 2.27$ (1.03), $t(558.51) = 2.20$, $p = 0.02$], however, it also determined that the same-sex male pairs and same-sex female pairs were not significantly different [$\beta = 1.24$ (0.98), $t(559.96) = 1.26$, $p = 0.21$].

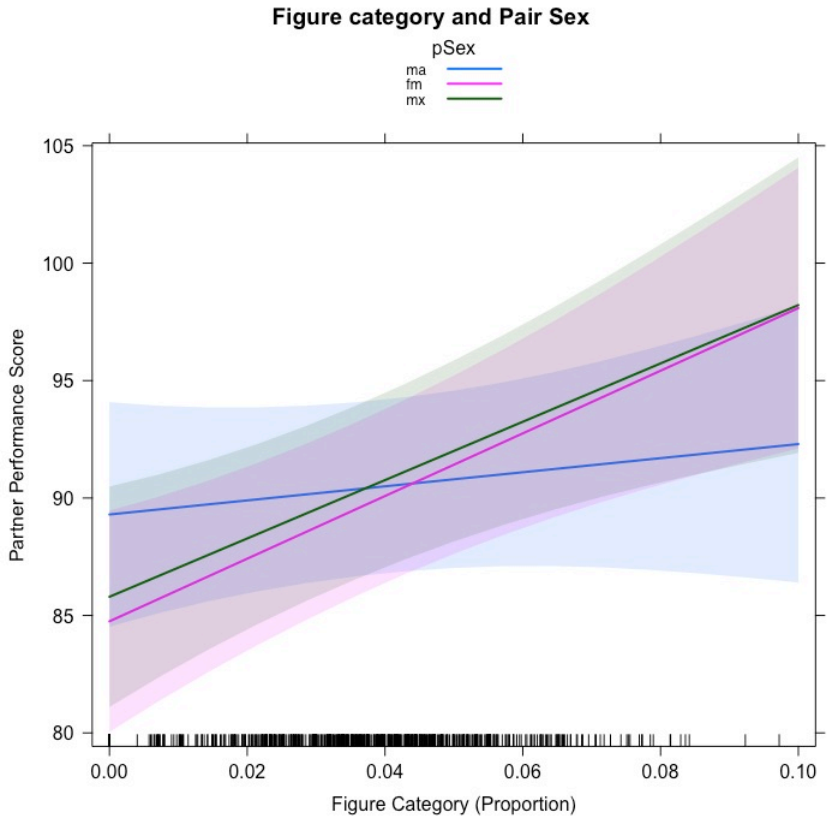

Figure 4. A plot of the variation across male pairs (ma), female pairs (fm), and mixed-sex pairs (mx) in the relationship between figure category usage and partner map performance. Fixed effects from the linear mixed-effects regression model plot relationships for male pairs in blue, female pairs in magenta, and mixed pairs in green. The shaded region represents a 95% confidence interval.

Figure 4 suggests that same-sex female pairs and mixed-sex pairs differed from same-sex male pairs. It appears that the relationship between partner performance and figure category usage was positive in female pairs and mixed-sex pairs. The results from mixed-effects regression including interaction terms between figure category use and pair sex confirmed that same-sex male pairs were not different from same-sex female pairs [β= 1.83 (0.97), t(565.17) = 1.88, p = 0.061], and they also did not differ significantly from mixed-sex pairs [β= 1.67 (1.00), t(561.09) = 1.67, p = 0.095]. The results indicated a marginal difference, as the p-values suggest.

Finally, a set of models examined whether talker sex alone moderated relationships between category usage and partner performance, regardless of pair type. Thus, linear mixed-effects regressions were carried out for each of the significant category types interacting with talker sex using with talker as the random intercept. Figure 5 below depicts the interaction between talker sex and figure terms use.

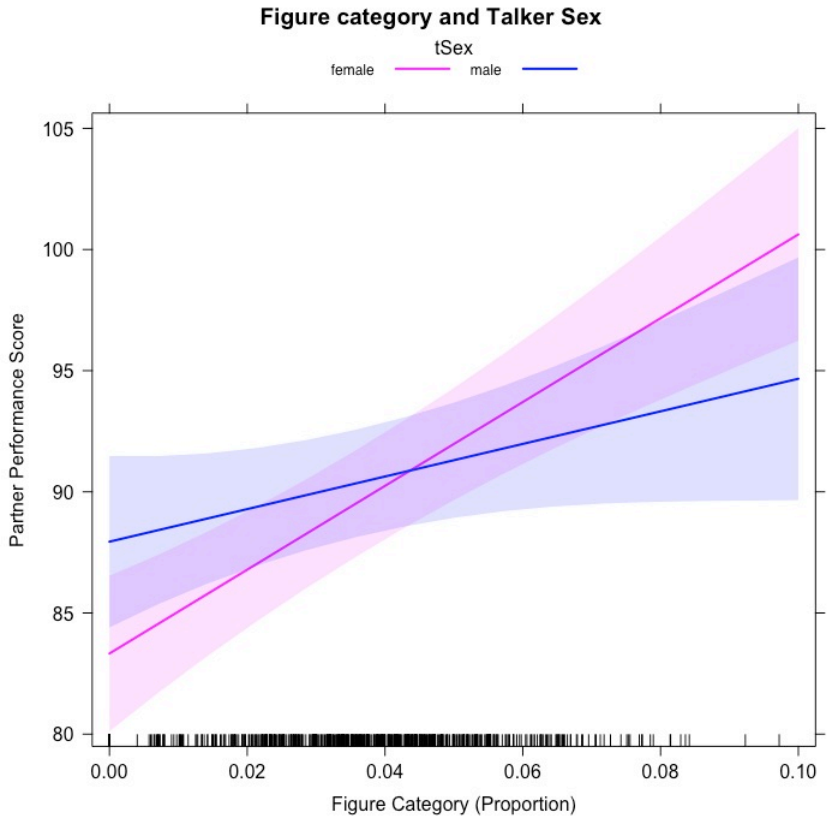

Figure 5. A plot of the variation across talker sex (male, female) in the relationship between figure category usage and partner map performance. Fixed effects from the linear-mixed effects regression model plot male talkers in blue and female talkers in magenta. The shaded regions indicate a 95% confidence interval.

As observed in Figure 5, the relationship between the use of the figure category and partner performance is stronger in female talkers than male talkers, regardless of their partner's sex. The mixed effects regression analysis revealed that, in fact, the relationship was significantly different between male talker and female talkers [β = -1.87 (0.86), t(571.27) = -2.172, p = 0.03]. Neither direction nor affirm category type produced any significant interactions with talker sex (p>0.05).

Similar to Pardo et al.'s (2019) finding that males produced a greater total number of words, same-sex male pairs produced a greater number of words within the context speech compared to same-sex female pairs and mixed-sex pairs. Similarly, male talkers produced more total number of words than female talkers. In addition, participants utilized some categories of words such as affirm and direction in a greater proportion than categories such as compare and figure. It was also revealed that the proportion of word usage differed significantly across various categories, but category usage did not vary across pair sex.

Relationships between performance and category usage suggested that the use of affirm, direction, and figure was significantly related to partner's performance. For instance, partner performance was negatively predicted by affirm category usage whereas it was positively predicted by the usage of direction and figure category. Further analysis revealed that the relationship between partner performance and individual category usage was weaker in male pairs when compared to female pairs and mixed-sex pairs. For instance, the use of affirm words negatively predicted partner's performance in female pairs and mixed-sex pairs whereas the relationship between performance and direction words was strongly positive in mixed-sex pairs. In other words, male pairs did not show significant relationships between partner performance and usage of either affirm, direction, or figure words. Further exploration revealed that male talker's use of figure words was positively correlated with partner performance, however, the relationship was stronger in female talkers.

Figure 6. A word cloud representation of words identified as constituents of the affirm category. In order from left to right, the individual clouds represent words produced within male pairs, female pairs, and mixed-sex pairs, respectively.

Male pairs showed no significant relationship between the use of affirm category words and partner performance. However, there are a few qualitative differences within the composition of affirm category words used by different pair sex, as seen in Figure 6. For instance, male pairs utilized 37 unique words, female pairs utilized a total of 53, and mixed-sex pairs utilized 35 unique affirm category words. Female pairs utilized a greater number of unique affirm category words in comparison to same-sex male pairs and mixed-sex pairs. Similarly, male pairs appear to utilize a greater number of words in higher frequency than female pairs and mixed-sex pairs. Words such as “okay” and “yeah” occurred on higher frequencies across all pair sex types whereas male pairs also utilized “alright” and “right” on a higher frequency.

Figure 7. A word cloud representation of words identified as the constituents of the direction category in order from the left to the right, male pairs, female pairs, and mixed-sex pairs, respectively.

Male pairs showed no significant relationship between the use of direction words and partner performance. However, male pairs produced 159 unique direction category words whereas female pairs and mixed-sex pairs produced 143 and 154 unique words each, respectively. It appears that the male pairs and the mixed-sex pairs produced a greater number of unique direction category words during their conversations, as evident in Figure 7. Words such as “to”, “on”, “right”, “down” and “above” appeared on higher frequencies across all pair sex.

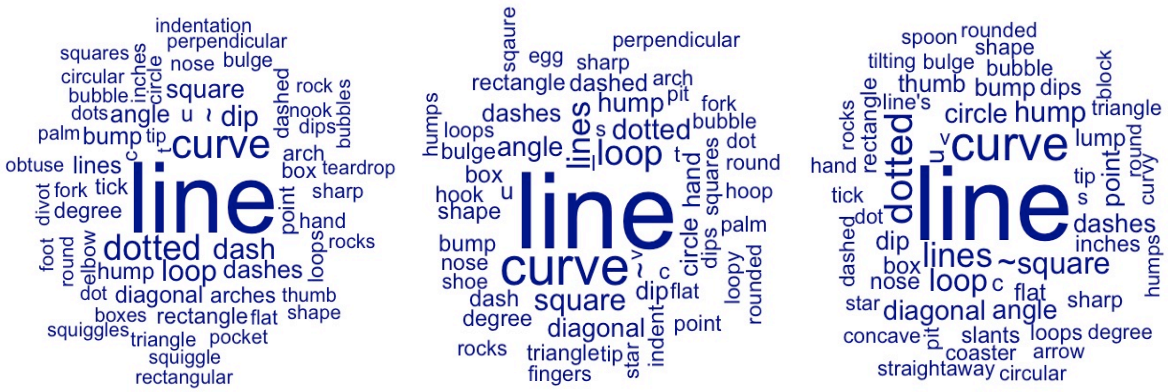

Figure 8. A word cloud representation of words identified as the constituents of the figure category within male pairs (left), female pairs (middle), and mixed-sex pairs (right), respectively.

For the figure category usage, the male pairs showed no significant relationship with their partner performance whereas female pairs and mixed-sex pairs were marginally significant. Male pairs utilized a total of 96 unique figure terms, female pairs utilized 106, and mixed-sex pairs utilized 101 unique words. The total number of unique figure terms utilized by pair sex types appears to be relatively similar. In addition, male pairs utilized words such as “dotted” and “dash” in a greater relative proportion than female pairs.

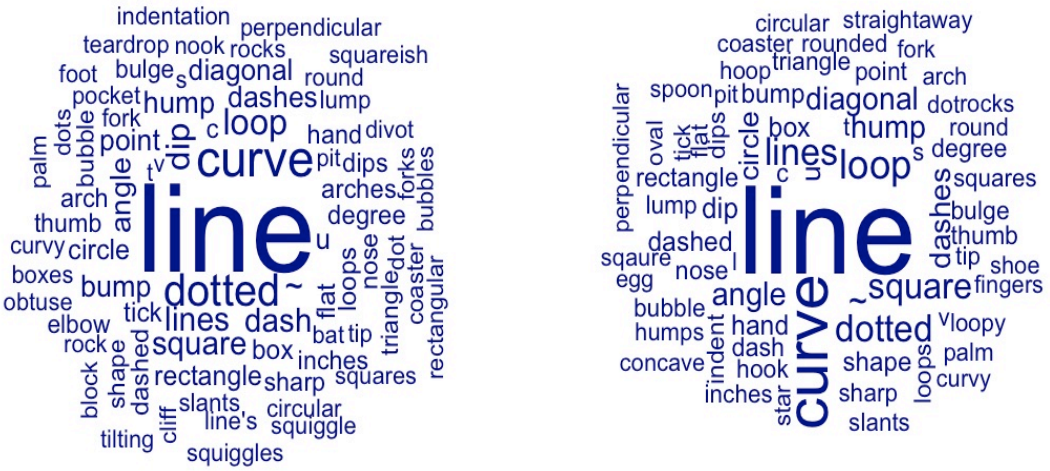

Figure 9. A word cloud representation of words identified as figure words in male talkers (left) and female talkers (right).

A further in-depth portrayal of the figure terms on the basis of talker sex is displayed in Figure 9. The relationship between the use of figure category words and partner performance was significant in female pairs. Male talkers produced a total of 121 unique figure category words whereas female talkers produced 128 unique words. It appears that the female pairs utilized a slightly greater number of unique figure terms in comparison to the male pairs. Words such as “line” and “loop” occurred in a greater frequency across both talker sex.

Discussion

The current study was aimed at exploring gender differences in conversational interaction, using a previously collected speech corpus by Pardo et al. (2019) to perform a word-level investigation of differences in communication style. Pardo et al. (2019) had found a negative relationship between the total number of words and partner performance was negative in male pairs and a positive relationship in female pairs. The Montclair Map Task Corpus was collected using a collaborative map matching task (Pardo et al., 2019), and the relationship

between task performance and various attributes of conversation styles such as the total number of words used was investigated. Similarly, Bell (2020) further examined gender differences in utterance goals and question-asking within the MMTC dialogues. The current study carried out a further in-depth analysis of the MMTC speech by coding context words into task-related categories (affirm, direction, figure, number, size, directive, compare, verb, function, pronoun).

Previous studies had suggested gender differences due to the demand characteristics of a context/setting (Leaper & Ayres, 2007). The Montclair Map Task is a task that required participants to navigate each other through routes within maps. Thus, it was hypothesized that categories such as direction and figure terms would be related to performance on the task, and the use of category would vary across pair sex and the relationship between category use and performance would also vary across pair sex. Our findings suggest that the direction terms and figure words played a vital role within the map task, and pair sex differences were observed in the relationship between their usage and partner performance. Contrary to the hypothesis, category usage did not vary across different pair sex.

The findings from Newman et al. (2008) that men use more spatial words such as "above" and "over" suggested that the words indicating direction would vary across male and female talkers. Also, across empirical studies and literature reviews by Mulac et al. (2001) and Mulac (2006), references to location and quantity were found to be male-linked features. Thus, the use of words that indicated information about the figure was also expected to vary across pair sex. Overall, the use of word categories could vary across different pair sex types and the use of different word categories could be related to partner performance.

Based on an analysis of words within context speech from Pardo et al. (2019), the current study found that male pairs produced more total number of words than female pairs and mixed-

sex pairs. Similar to the findings of Pardo et al. (2019), male talkers produced more words within context speech than female talkers. Since the context speech comprised a majority of corpus speech, the findings from Pardo et al. (2019) were further confirmed. An in-depth analysis of word categories within the context speech revealed that the proportion of word usage varied significantly across various categories. In other words, some categories were utilized in greater proportion than other categories. For instance, direction category words were used in a greater proportion in comparison to size category words. The Montclair Map Task requires the use of spatial terms and words that indicate a direction, thus, a higher proportion of direction category words might be utilized.

In addition, the use of only affirm, direction, and figure category words was found to be significantly related to partner performance. This confirmed our assumption that partner performance varied with the use of different kinds of words. The use of affirm category negatively predicted partner performance whereas the use of direction and figure terms positively predicted partner performance. It is possible that the use of direction indicative words assisted with the map-matching task, and the use of figure terms also contributed towards the same goal. In particular, the relationships between performance and category usage were found to be stronger in female pairs and mixed-sex pairs. For instance, the relationship between the use of affirm words and partner performance was found to be stronger in female pairs and mixed-sex pairs while it was also negative. Leaper and Ayres' (2007) found that women utilize affiliative language features such as language that denotes agreement. However, it is possible that the affirm words are not instrumental to the map task. Thus, female talkers' use of these words in female pairs and mixed-sex pairs could negatively impact task performance.

Similarly, the relationship between the use of direction category and partner performance was found to be stronger in mixed-sex pairs, and the relationship was positive. According to Mulac et al. (2001) and Mulac (2006), references to location and quantity were utilized more by men, and Newman et al. (2008) also found that men used more spatial terms. While Mulac (2006) transformed language features count into a percentage of the transcript, Mulac et al. (2001) had reported a number of studies that identified these language features as male-linked. On the other hand, Newman et al. (2008) utilized frequency information. The current study utilized the proportion of words in order to eliminate the effect of a greater total number of words produced by male talkers. It allowed for a comparison of the relative usage of different kinds of words as predictors of performance, regardless of the total number of words. Thus, the findings from the current study dismiss the possible relationship between partner performance and direction category usage in male pairs. However, the findings were significant for female pairs and mixed-sex pairs. A possible explanation could be that same-sex male pairs did not produce a composition of direction words that was similar to that produced by female pairs and mixed-sex pairs. As seen in Figure 7, there are, indeed, some qualitative differences such as a somewhat greater variety of direction words between pair types.

Again, the relationship between the use of figure category and task performance was found to be stronger in female pairs and mixed-sex pairs. Since the use of figure category positively predicted partner performance, it can be assumed that a lack of significant relationship between the category usage and task performance in same-sex male pairs could indicate that the male pairs' performance depended on factors other than figure terms.

In other words, male pairs did not show significant relationships between partner performance and usage of either affirm, direction, or figure words. Further exploration revealed

that the male talker's use of figure words was weakly correlated with partner performance whereas the female talker's use of figure words was strongly correlated with partner performance, regardless of pair type. The current findings shed light on the relationships between word category usage and partner performance. Perhaps, the social constructionist/contextualist approach to explain gender differences as the response to demand characteristics of a context/setting might explain the current findings (Leaper & Ayres, 2007). In other words, the differences in the relationship between affirm, direction, and figure category usage and performance across different pair sex are due to the map-task setting. For female pairs and mixed-sex pairs, the findings supported the current hypotheses. In contrast, the relationships between significant categories such as affirm, direction, and figure and task performance were not significant in same-sex male pairs. These findings indicate that the male pairs' performance was associated with factors other than the category of words.

Newman et al. (2008) had utilized the frequencies of word categories to obtain differences in the usage of different types of terms across gender. Conversely, Park et al. (2016) calculated the relative use of each topic for every user, where topic use was defined as the probability of using a topic. On the other hand, the current study was based on the proportion usage of different types of words. It allowed for a comparison of the relative usage of different kinds of words as predictors of performance, regardless of the total number of words. In addition, it allowed for further examination of the previous finding of Pardo et al. (2019) that males used more words, and that was negatively related to performance by determining whether relative usage of specific kinds of words predicted performance. Since the current study did not find any significant differences between pair types in their overall proportion usage of different categories of words (i.e., nonsignificant interaction between pair sex and category detailed in Appendix C),

previous findings that males utilized direction and figure related words and instrumental language such as size, number, and directives whereas females utilized affirming words, pronouns, and compare words were not validated in this corpus. Proportion usage is a better metric for comparison because findings of a greater frequency of usage of a particular category would not be meaningful when the baseline number of words differed across categories.

No prior studies have investigated word category usage across pair sex types in task-oriented conversational interaction. In addition, there is a lack of evidence within the literature to support any relationships between the use of a certain category of words and performance on a task. In other words, studies have not yet examined word category usage patterns and task performance. In that respect, the current study is a novel endeavor and the current findings are novel findings. A recent study can provide some insight into the current findings (Koulouri et al., 2017).

Koulouri et al. (2017) studied gender differences in communication in complex, collaborative tasks and how it influences task performance. The task involved a participant in the role of an instructor guiding another participant who was designated as a follower to six locations in a simulated town. The experiment included same-sex and mixed-sex gender pairs. As a Computer-Supported Collaborative Work (CSCW), visual feedback about objects and events was provided to the participants during the task. The experiment also excluded gender-related social preconceptions by instructing the participants that they were interacting with a robot, which is informed by the social constructionist approach to understanding gender differences in communication (Leaper and Ayres, 2007). Koulouri et al. (2017) did not find any performance difference between males and females, however, differences were observed in their communication styles. Males tended to utilize novel vocabulary in the event of a communication

problem whereas females resorted to a more familiar speech behavior. Especially in the absence of visual feedback, females increased the quality and specificity of their interactions, and their strategies such as repeating and re-using vocabulary were effective. However, males continued to present novel vocabulary when communication problems occurred and the visual feedback was also absent. Koulouri et al. (2017) is among a handful of studies that associated language features with performance in an objective task.

Perhaps, Koulouri et al.'s (2017) findings might explain the current findings. Males utilized novel vocabulary to overcome miscommunication whereas females utilized a more familiar set of speech, and that females utilized rich verbal means to compensate for the lack of visual cues (Koulouri et al., 2017). Similarly, females increased the specificity of their interactions and provided elaborate, detailed verbal contributions whereas males resorted to explorative behavior such as using novel words. More importantly, females' strategies were found to be more effective than males' when visual feedback was removed.

On the other hand, the Montclair Map Task was essentially a collaborative task without any visual feedback, therefore, it is possible that females utilized more effective strategies to overcome the lack of any visual feedback by deploying linguistic tools. For instance, within the MMTC, male pairs produced a slightly greater variety of direction and figure category words in comparison with female pairs however, the relationship between the category usage and partner performance was not significant. Perhaps the females utilized more effective strategies such as repeating and re-using the vocabulary instead of utilizing novel words. Thus, same-sex female pairs and mixed-sex pairs could be showing a stronger relationship between the category use and partner performance due to conservative strategies whereas males' use of novel vocabulary impacted these relationships negatively.

Although Koulouri et al. (2017) utilized a computer-mediated setting, their findings apply within the task-oriented lab setting of the Montclair Map task. When the visual feedback was removed, the navigation task in Koulouri et al. (2017) resembled the Montclair Map Task since the participants were unable to see each other, and they did not receive any visual feedback. Participants in Koulouri et al. (2017) were told that they were interacting with a robot, and they were completing the task on a computer whereas participants in the current study completed the task on a paper by interacting with the knowledge of interacting with another person of either the same or different gender. Since there are no prior studies examining the proportion usage of a word category and task performance, evidence from Koulouri et al. (2017) supports our current findings that the relationships between the use of affirm, direction, and figure words and performance varied across different gender pairs.

The current study had a few limitations that need to be discussed. Although the word categories were formulated using previous research and task-specific context, it is possible that the categories were not as clearly delineated as assumed. For instance, the word "right" could be correctly categorized as either a direction category word or an affirm category word. In order to resolve this issue, the current study made use of a proxy category, *ambi*, which was then further categorized by examining the context of each usage. However, there could be many instances of such ambiguous words that were categorized inaccurately. Therefore, a bag-of-words approach to text analysis appears to be inadequate for the purpose of investigating the meaning of words. Future studies should utilize semantic parsing instead of the bag-of-words model in order to produce more accurate word categories. Semantic parsing is feature-rich since it relies on the individual word and its word order i.e. a word is considered in relation to the words occurring before and after it (Fillmore, 2012).

On the other hand, most words were categorized into one of twelve categories, and only the words that were not relevant to the task or that did not appropriately match any other categories were coded into the *other* category. This ensured that the semantically valuable words did not get excluded from the analyses. In addition, some context speech also included either partial or complete landmark label phrases, which were delineated into a *landmark* category.

Although we did not find gender differences in the proportion usage of different types of word categories, the current study found that affirm, direction, and figure categories were associated with the partner performance. The relationship between the proportion usage of these categories and the performance of a partner on the map task was weaker in male pairs while it was stronger in female and mixed-sex pairs. Same-sex male pairs were found to be different from female and mixed-sex pairs in the original MMTC analyses of word counts and relationship with performance (Pardo et al., 2019). The current study supported the expectation that the male pairs would differ from female pairs and mixed-sex pairs in the relationship between category usage and performance. However, expectations that the category proportions would vary across different gender pairs were not supported by the current study. Further analyses of the Montclair Map Task Corpus should be conducted using semantic parsing to provide additional accuracy to the categorization, and detailed qualitative analyses of the words utilized by male pairs need to be assessed. Ultimately, the current study found some meaningful gender differences that further supported and explained the previous analyses of the Montclair Map Task Corpus, and provided a promising venture for future study.

References

Andersen, P. A. (2006). The evolution of biological sex differences in communication. In D. J. Canary & K. Dindia (2nd Eds.), *Sex Differences and Similarities in Communication* (2nd ed., pp. 117-135). Mahwah, NJ: Lawrence Erlbaum.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weiniert, R. (1991). The HCRC Map Task Corpus. *Language & Speech*, *34*(4), 351–366.

Anderson, K. J., & Leaper, C. (1998). Meta-analyses of gender effects on conversational interruption: Who, what, when, where, and how. *Sex Roles*, (3–4), 225-252.

Bell, C. (2020). *Question-Asking in conversational tasks: A gender comparison* [Unpublished master's thesis]. Montclair State University.

Fillmore, C. J. (2012). Encounters with Language. *Computational Linguistics*, *38*(4), 701-718.

Giles, H., Taylor, D. M., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: some Canadian data. *Language in Society, 2(02), 177.* doi:10.1017/s0047404500000701

Giles, H., Coupland, N., & Coupland, J. (1991). *Accommodation theory: Communication, context, and consequence.* In H. Giles, J. Coupland, & N. Coupland (Eds.), *Studies in emotion and social interaction. Contexts of accommodation: Developments in applied sociolinguistics* (p. 1–68). Cambridge University Press; Editions de la Maison des Sciences de l'Homme. https://doi.org/10.1017/CBO9780511663673.001

Hall, J. A. (2006). Nonverbal behavior, status, and gender: How do we understand their relations? *Psychology of Women Quarterly*, 30, 384-391.

Hartman, M. (1976). A descriptive study of the language of men and women born in Maine around 1900 as it reflects the Lakoff hypotheses in language and woman's place. In B. L. Dubois & I. Crouch (Eds.), *The sociology of the languages of American women* (pp. 81–90). San Antonio, TX: Trinity University Press.

Koulouri, T., Lauria, S., & Macredie, R. (2017). The influence of visual feedback and gender dynamics on performance, perception and communication strategies in CSCW. *International Journal Of Human-Computer Studies*, *97*, 162-181. doi: 10.1016/j.ijhcs.2016.09.003

Lakoff, R. (1975). *Language and woman's place*. New York: Harper Colophon Books.

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.

Maltz, D. N., & Borker, R. A. (1982). A cultural approach to male-female miscommunication. In J. J. Gumperz (Ed.), *Language and Social Identity* (pp. 196-216). Cambridge, UK: Cambridge University.

Mulac, A., Bradac, J. J., & Gibbons, P. (2001). Empirical support for the gender-as-culture hypothesis. *Human Communication Research*, 27(1):121–52.

Mulac, A. (2006). The gender-linked language effect: Do language differences really make a difference?. *Lawrence Erlbaum Associates Publishers*.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes* ;45(3):211–36.

Pardo, J., Urmanche, A., Gash, H., Wiener, J., Mason, N., Wilman, S., Francis, K., & Decker, A. (2019). The Montclair Map Task: Balance, Efficacy, and Efficiency in Conversational Interaction. *Language And Speech*, *62*(2), 378-398. https://doi.org/10.1177/0023830918775435

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Park, G., Yaden, D. B., Schwartz, H. A., Kern, M. L., Eichstaedt, J. C., Kosinski, M., Stillwell, D., Ungar, L. H., & Seligman, M. E. P. (2016). Women are Warmer but No Less Assertive than Men: Gender and Language on Facebook. *PLOS ONE, 11*(5), e0155885.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013) Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS one, 8*(9). https://doi.org/10.1371/journal.pone.0073791

Appendix A: Sample MMT map pair

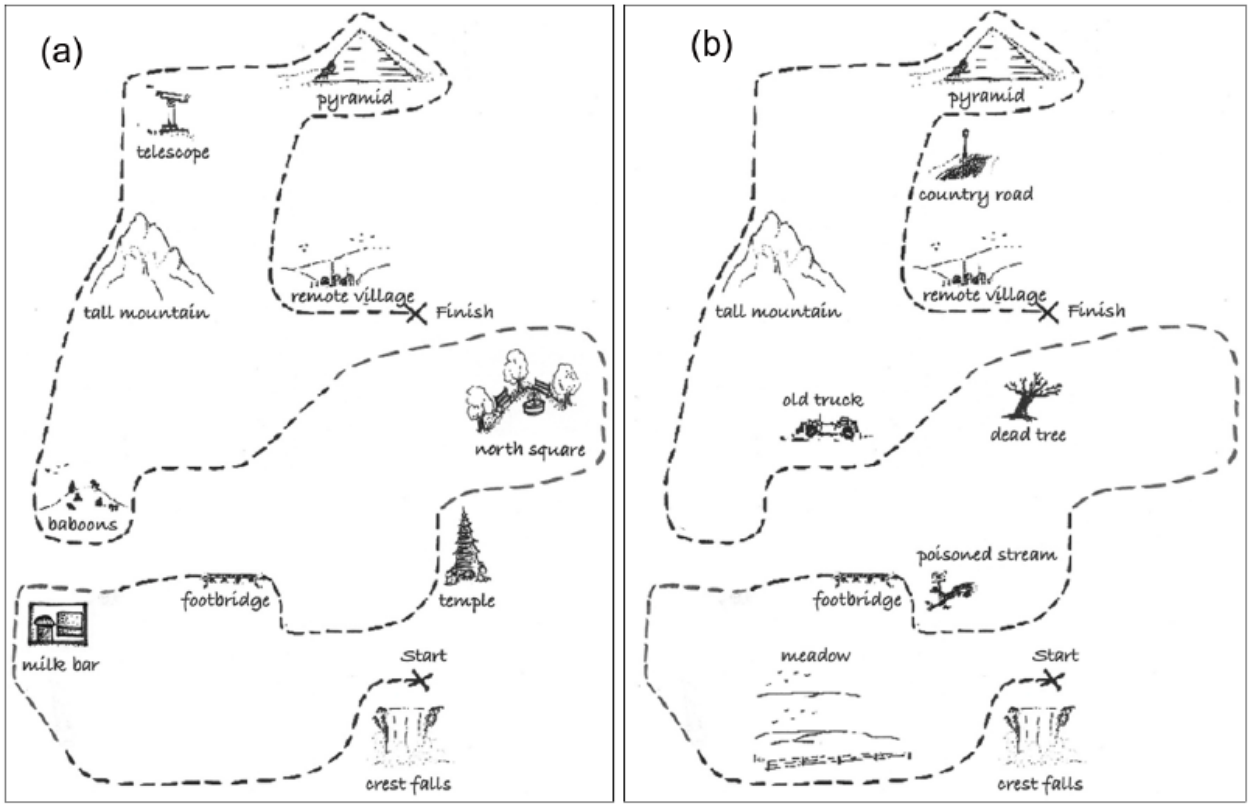

Figure 1: A single map pair from MMT

Appendix B: Excerpt from Task Transcriptions

Talker A: from the start um there would be the camera shop

Talker B: yeah I have that one too

Talker A: and then under that is the parked van

Talker B: that one I'm missing

Talker A: there's a museum okay and then where that line go --slopes down is the yacht club

Talker B: okay I have that one too

Appendix C: Table Results

Table 1.

2-way Repeated Measures ANOVA between pair sex and category of words

Predictor	df_{Num}	df_{Den}	*Epsilon*	*F*	*p*	η^2_g
pSex	2.00	93.00		0.55	.580	.00
Category	4.03	374.55	0.45	1079.47	.000	.92
pSex x Category	8.05	374.55	0.45	1.29	.247	.03

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, *p*-values and degrees of freedom in the table incorporate this correction. η^2_g indicates generalized eta-squared.

Table 2.

Linear Mixed Effects Regression Results

			Partner's Map Task Performance		
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*df*
(Intercept)	90.23 ***	0.77	88.71 – 91.75	117.77	92.36
affirm	**-1.15 ***	**0.50**	**-2.13 – -0.17**	**-2.31**	**537.86**
compare	-0.14	0.50	-1.12 – 0.84	-0.28	535.40
direction	**1.32 ***	**0.54**	**0.26 – 2.39**	**2.44**	**564.60**
directive	0.62	0.40	-0.17 – 1.41	1.54	553.28
figure	**2.05 ***	**0.47**	**1.13 – 2.98**	**4.36**	**565.48**
number	0.22	0.42	-0.61 – 1.05	0.52	563.13
pronoun	0.19	0.51	-0.80 – 1.19	0.39	565.99
size	-0.11	0.41	-0.93 – 0.70	-0.28	560.56
verb	-0.46	0.52	-1.48 – 0.55	-0.90	564.93
				** p<0.05* *** p<0.01* **** p<0.001*	

Linear Mixed Effects Regression model for category of word type predicting partner's map task performance. Confidence intervals are set at 95%.

Model Formula: Performance score ~ scale(affirm) + scale (compare) + scale(direction) + scale(directive) + scale(figure) + scale(number) + scale(pronoun) + scale(size) + scale(verb) + (1|Talker)