2019

The Development of a Lexicon for the Communication of Action in Cooperative Work

Claire Supriya Shah
Wright State University

THE DEVELOPMENT OF A LEXICON FOR THE COMMUNICATION OF ACTION IN COOPERATIVE WORK

A thesis submitted in partial fulfillment of the
requirements for the degree of
Master of Science

By

CLAIRE SUPRIYA SHAH

B.S., Towson University, 2014

2019
Wright State University

WRIGHT STATE UNIVERSITY
GRADUATE SCHOOL

August 28, 2019

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Claire Supriya Shah ENTITLED The Development of a Lexicon for the Communication of Action in Cooperative Work BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

_____
Valerie L. Shalin, PhD
Thesis Director

_____
Debra Steele-Johnson, PhD
Chair, Psychology

Committee on Final Examination:

_____
Scott Watamaniuk, PhD

_____
Ion Juvina, PhD

_____
Barry Milligan, Ph.D.
Interim Dean of the Graduate School

ABSTRACT

Shah, Claire Supriya. M.S. Department of Psychology, 2019. The Development Of A Lexicon For The Communication Of Action In Cooperative Work.

This research expands upon the research conducted by Clark and Wilkes-Gibbs (1986) on how individuals collaborate and reach common ground in the domain of objects into the domain of action. Pairs of participants (N = 22) were asked to complete a set of six maneuvers with a remote-control car. Dialogue was transcribed and analyzed for total word count, verb phrase count, number of turns taken, number of errors committed, and selected other linguistic characteristics. Total word count, verb phrase count, number of turns taken, and number of errors committed all significantly decreased over time, either linearly or logarithmically. This research shows support for a general distinction between path and manner verbs by showing different associated language patterns for the different verb types. A key finding in this study is that learning of path maneuvers is dependent on learning features in the environment, whereas this is not important in manner maneuvers.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

I. INTRODUCTION

As individuals learn a new task and attempt to communicate, they create their own jargon to describe task elements. Clark and Wilkes-Gibbs (1986) showed that novices use fewer and more concrete noun phrases in collaborative communication while learning to sequence a set of abstract objects. Their simple sequencing task avoided the need for verbs. Nevertheless, linguistic research has suggested that grammar rules are the same for noun and verb phrases (Chomsky, 1953). Therefore, similar language patterns should emerge in both noun and verb phrases. Switching from the domain of objects to the domain of action makes two contributions. First, changes in the language to describe complex, previously unfamiliar action sheds light on the development of schemas. Second, from a practical perspective, an operational definition of schemas grounded in their development can facilitate the development of tools to support planning and documentation. Thus, the purpose of my study is to identify patterns in verb phrases used in cooperative communication while planning to execute complex tasks.

Schemas

Schemas reflect organized patterns of thought or action that specify temporal and spatial relationships among individual components. Bartlett (1932) believed that schemas result from sociocultural influence, demonstrated in his War of the Ghost experiments. Bartlett (1932) asked participants to read the Native American folktale "War of the Ghost" and recall as much as they could from the story in several time intervals up to one year later. The results showed that participants recalled information

that matched their own culture, and omitted information that did not align. Participants also modified the content that they remembered to match their own

socially grounded schemas. Both findings support the claim by modifying detail, schemas reflect sociocultural influence, a kind of tacit agreement about the relevant content in such stories.

However, subsequent research in cognitive psychology emphasized the apprehension of schemas as a reflection of individual cognitive capability, describing for example how verbal description could result in procedural knowledge, eventually consolidated to reduce the reflection needed to respond in a task environment (Anderson, 1977). Nevertheless, anthropologists continued to emphasize the cultural explanation for the origins of schemas, identifying their role in the coordination of distributed work. Livingston (1987) defined accepted work methods as agreed upon work practices of a discipline. Accepted methods help team members predict other participant's behavior and assist in monitoring for errors (Shalin, Geddes, Bertram, Szczepkowski, & DuBois, 1997).

Accepted methods may also influence the level of description in documentation and dictate the units of analysis for planning and hence planning tools. Many application domains appear to depend on accepted methods, such as medicine, military operations, and space exploration. Yet task and work analysis lack principled procedures for identifying the contents, scope and language of accepted methods.

Consistent with Clark and Marshall, (1981) accepted methods provide common ground in an established community of practice. Common ground is an important conversational tool that is defined as "mutual knowledge, beliefs, and assumptions" (Isaacs & Clark 1987; Clark & Marshall, 1981). We hypothesize that over time, accepted methods become associated with a shared lexicon. This shared lexicon facilitates the distribution of information throughout the workplace, such as instructions. The practical motivation for developing a shared lexicon includes efficient and effective information exchange in planning.

Building on the work proposed by Newman, the present study examines changes in language with task experience as an indicator of the development of accepted methods. Newman's proposed work consists of an observational study in the domain of Martian surface exploration (M. Newman, personal communication, September 2015). While Newman's study provides face validity, numerous complications arise in interpreting the resulting data. First, there was just one team (n=1) observed. Second, uncontrolled external events, such as team member rotation and equipment malfunctions influenced change in language. These complicate the recovery of a function that represents language changes over time. To address these complications in the interpretation of observational data, the present study will examine language changes for multiple dyads in an experimental repeated action-sequencing task. I apply an established paradigm from psycholinguistics, originally developed for a task sequencing objects, to a sequence of actions with a remote-controlled car.

Referential Language

Clark and Wilkes-Gibbs (1986) found that as a pair of individuals work cooperatively to complete a task, the amount of language necessary to identify an object decreases. In their study, participants arranged tangram images in a pre-specified order. Each participant served as either the director or the matcher. The director had a representation of the required order of tangrams and directed the matcher on the desired order of the tangram images. In each trial, the order of the tangram images was randomized, and the participants completed six trials. Tangrams were chosen because of the perceived abstract nature of the objects. Clark and Wilkes-Gibbs (1986) recorded the time it took participants to complete the task, and transcribed all spoken communication including "changes of speaker, back-channel responses, parenthetical remarks, interruptions, hesitations, false starts, and basic intonational features" (p. 11). Although the authors do not report results for all of the components transcribed, they found significant declines in changes of speaker, time on task, total word count, and noun phrase count. They explained their results using the concept of "common ground" developed through communication regarding shared work goals. The development of a shared lexicon reflects common ground by enabling the reduction of communication, in turn reducing time on task. As Clark and Wilkes-Gibbs (1986) showed, the number of noun phrases used in the cooperative communication between participant A and participant B in his tangram experiment decreased after each trial. Linguistic theory

states that noun and verb phrases are governed by the same underlying structure (Chomsky, 1970), predicting the same changes in language patterns for the sequencing of previously unfamiliar action as objects. This linguistic theory leads me to my hypotheses:

Hypothesis 1: Verb phrase count will decline as participants progress in the planning process.

Hypothesis 2: The total number of words used by participants will decline as participants repeat trials of sequencing.

Action-based tasks introduce complexity that Clarke and Wilkes-Gibbs did not encounter. The complexity has the potential to cause participants to commit more errors than in the object-based task. To account for this, I will analyze task accuracy over trial progression.

Hypothesis 3: Task accuracy will increase as participants progress through repeated trials.

Hypothesis 4: As the participant pairs progress through the trials, they will assign mutually agreed upon terms to the action tasks of the remote-controlled car.

Hypothesis 5: The time needed to complete the string of 6 tasks will decrease as the participants progress through the planning process.

Principled selection of action stimuli

The type of verb potentially impacts lexical selection and word count. English verbs distinguish between manner but not generally path. For example, separate words

distinguish walking from running and skipping. Path indication generally requires a prepositional phrase, e.g., walk into, run by, etc. Exceptions include enter and exit. I note additionally that prepositions require arguments, that is grammatical objects. We walk into *the house* or run by *the lake*. The need for arguments in the specification of path anticipates familiarity with the task environment. Reliance on the prepositional phrase in English places a constraint on the reduction of word count, and more generally illustrates the need for a principled selection of action stimuli.

Several linguists have attempted to categorize actions in order to link semantics with syntactic requirements on sentences (Jackendoff, 1991; Talmy, 2000). Roger Schank theorized that verbs can be categorized into 11 conceptual primitives: ATRANS, ATTEND, INGEST, EXPEL, GRASP, MBUILD, MTRANS, MOVE, PROPEL, PTRANS, SPEAK (1972). Five of these primitives describe physical actions studied here: INGEST (take something into an animate object), EXPEL (take something from inside and force it out), GRASP (physically grasp something), MOVE (move a body part), PROPEL (apply force to an object). State change primitive actions include PTRANS (change physical location of an object) and ATRANS (change an abstract relationship of a physical object). Mental acts include MTRANS (transfer information mentally) and MBUILD (combine or create thoughts). Other primitives include SPEAK (make a sound) and ATTEND (direct a sense organ or focus towards a stimulus) (Schank & Abelson, 1988). Ideally, these conceptual primitives in actions can represent any sentence with any structure (Schank, 1972).

Conceptual dependency theory specifies the arguments for action with implications for word count. These argument categories are: PP (physical object), ACT (one of the above eleven primitive actions), LOC (location), T (time), AA (modifications of aspects of an ACT), and PA (attributes of an object).

However, conceptual dependency theory does not establish the category of particular words. Levin (1993) provides an inventory for English verbs, associating them with categories generally consistent with conceptual dependency theory: Verbs of Sending and Carrying, Verbs of Change of Possession, Verbs of Contact by Impact, Poke Verbs, Verbs of Perception, Verbs of Social Interaction, Verbs Involving the Body, Verbs of Lingering and Rushing (Levin, 1993).

Specifically, I sampled from PROPEL (manner) and PTRANS (path) related actions for this study. These conceptual primitives aided in the classification and distinction between path, manner, and combination maneuvers specified in the following section. PROPEL variants of manner may be more amenable to single word capture while PTRANS variants may persist as multi-word noun phrases due to the articulation of location. A final issue in the selection of experimental stimuli concerns the relationship between the elements in the set. Stimulus similarity determines which features are diagnostic and therefore notable. For example, a set exclusively composed of PTRANS may necessitate persisting location referents whereas a set of a single PTRANS in the context of exclusively PROPEL may obviate the need to specify location in PTRANS.

Linguistic characteristics. Dialogue can be classified in a multitude of different ways, as shown above. To analyze the differences between verb types properly, the present study uses a categorization scheme developed for the Linguistic Inquiry and Word Count software (Pennebaker et. al., 2015). Of the approximately 90 categories developed for this software, I have selected 19 to focus my analysis. The categories for word count and common verbs directly identify measures to conduct the replication analyses of Clark and Wilkes-Gibbs' (1986) original study. To ensure all verb phrases were counted, the categories for auxiliary verbs and common adverbs were also analyzed. The other parts of speech I chose to analyze are prepositions, interrogatives, and comparatives. The overarching cognitive processes category was chosen to identify the underlying thought processes of the chosen language. This category consists of insight words, causations, discrepancies, tentativeness, and certainties (Pennebaker et. al., 2015) perhaps best associated with metacognitive processes. The informal speech category of assent words will be used to analyze the backchannel responses similarly to the original study (Clark & Wilkes-Gibbs, 1986). The overarching time orientation category was chosen to further investigate the action orientation of the stimuli. This includes motion, space, time, past focused words, present focused words, and future focused words. These analyses are purely exploratory; therefore, I do not have any related hypotheses.

Method

Participants

Fifty undergraduate-level students (25 pairs) attending Wright State University in Dayton, Ohio, who are enrolled in psychology classes participated in this study. Four pairs failed to complete the study. Data from ten pairs were discarded due to extenuating factors including not meeting English requirements as stated in the recruitment requirements and technical malfunctions. Data from the remaining 22 participants (11 pairs) are used in this study[1]. Students received course credit through an online sign up system (SONA) for their specific class requirement.

Demographics. A demographics measure was administered to the participants. The demographics measure is included in Appendix A. Participants included 59% female and 41% male with an average age of 22.41 years old (M = 20.76 removing outlier of 59 years old). Ten participants identified as white, ten participants identified as black, one participant identified as Asian, and one participant identified as Hispanic. The questionnaire also included a question asking if the participant had experience with remote controlled cars. Thirteen out of the twenty-two participants responded "yes", eight responded "no" and one left the question blank.

[1] The number of participants and trials is comparable to Clark and Wilkes-Gibbs (1986), suggesting sufficient power. Twenty-two participants (11 pairs) participated in the study, which is greater than the number of participants in Clark and Wilkes-Gibbs (1986). We slightly increased the participants because the mean time taken to complete the task was predicted to increase, which in turn increased the variance.

Equipment

The remote-control car used was a Tera WLtoys A999 1:24 Electric 2WD Remote Control RC car. A set of maneuver videos were pre-recorded in the experiment room, from the perspective of the driver, displayed on a first-generation iPad Air. The maneuver sequences for trials two through five were randomized using a random number generator and the maneuver clips were edited together using iMovie. The first and sixth maneuver sequences were consistent throughout all participant pairs. The first maneuver sequence was formed by alternating the different types of maneuvers: path1, combination1, manner1, path2, combination2, manner2. This was done to minimize the amount of carryover language from one maneuver to another. The sixth maneuver sequence was formed by taking the first sequence and reversing the order of the two sets of three: manner1, combination1, path1, manner2, combination2, path2.

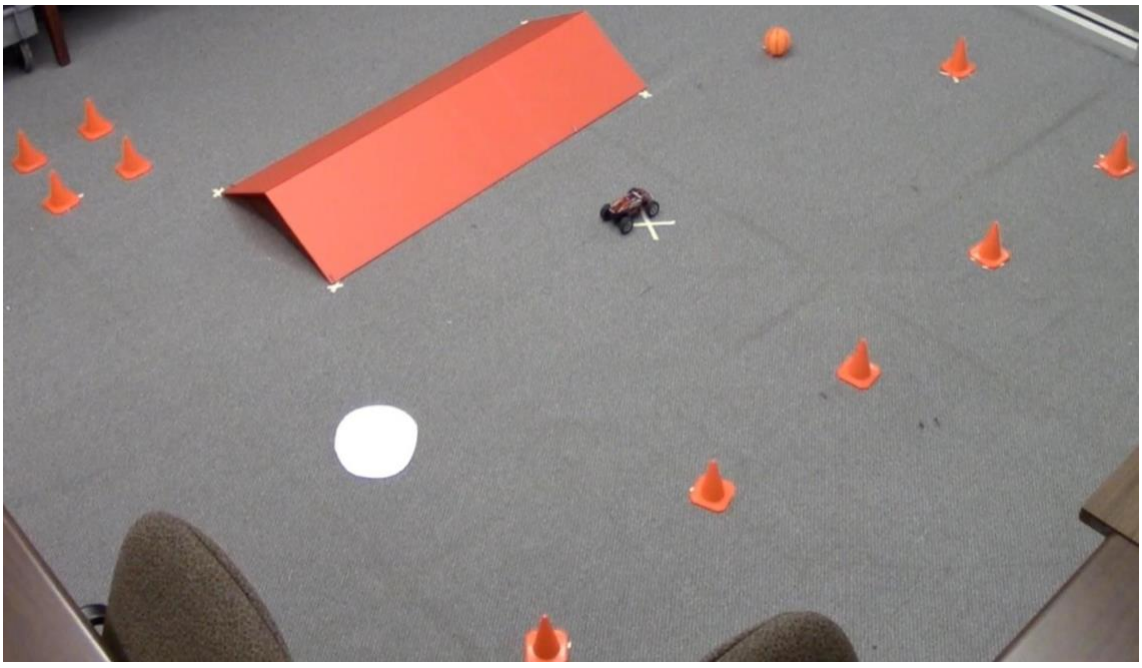

Figure 1: Remote control car in the experimental set up.

Stimuli

Participants executed a series of six maneuvers in each trial using a remote-control car within an obstacle course. Trial number (one through six) acted as time for my independent variable[2]. The six maneuvers chosen are listed below divided into categories based on verb type with their description and error criteria.

List of Maneuvers and Corresponding Error Allowances

1. Path Maneuvers

[2] I started with a list of 16 maneuvers and conducted pilot testing with graduate students and solicited feedback to identify the maneuvers that were feasible in a two-hour timeframe. The two-hour timeframe was selected to reduce participant fatigue.

a. Place front right tire onto an object on the floor (paper circle) – must place only the front right tire on the paper circle. It is incorrect if the driver places any other part on the circle.
b. Drive the car in 1.5 circles – must start in the same spot as the video and have to end at the same spot within 2 inches

2. Manner Maneuvers
    a. Move the car back and forth 8 times in short bursts – short bursts are about 6 inches in length
    b. Lightly tap the front of the car on the back wall – must approach the wall slowly, tap can not make a sound
3. Path and Manner Maneuvers
    a. Reverse in a straight line against the ramp – must begin at the top right corner of the ramp and end on the bottom right corner.
    b. Reverse into a "parking space" – must reverse into the cones and follow the same pattern as the video.

Task

Participants in each dyad were randomly assigned into either the role of director or driver. The director watched video footage of the remote-control car executing a series of maneuvers and then directed the driver on how to execute the actions correctly. The driver used the instructions from the director to execute the sequence of

maneuvers correctly. After each maneuver, the car returned to the center of the room on a white X. This allowed independent analysis of the language used for each maneuver.

I created standard first and last maneuver sequences to ensure comparability between participant pairs. The first and last sequences were the same across all participant pairs and were created with the intent to minimize language transfer between maneuvers. For trials two through five, the maneuver orders were randomized for each participant pair. This maximized independence of the common ground formed within the sequence of maneuvers from the maneuvers themselves.

Measures

Process. All trials were recorded (video and audio). All of the verbal communication was personally transcribed by the experimenter. Each participant pair's conversation was transcribed into Word documents by the experimenter and then analyzed. The conversations between the two participants were then run through Language Inquiry and Word Count (LIWC) software to determine the linguistic characteristics of the dialogue (Pennebaker et al. 2015). LIWC produces an output of proportions of the different types Word count was measured using the LIWC software as integers. Proportions enable content analysis independent of total word count, as we predicted total word count would significantly decrease over trials.

Specific categories of interest were prepositions, auxiliary verbs, adverbs, compare, insight, discrepancy, verbs, tentative, certain, focus on past, focus on present, focus on future, motion, space, time, assent, and interrogatives. LIWC outputs the

proportion of the words in the category that were found in the transcriptions. Verb phrases were identified by the proportion of verbs given in the output from LIWC.

Outcome. Turn-taking was defined as the number of times the speaker switched between the driver and director. Any communication switches with the experimenter were excluded from this measure. Errors for each maneuver were determined during the experiment using the corresponding error allowances described in the task section within the method section. When a participant pair performed a maneuver outside of the error allowance, the experimenter stated "Incorrect" which informed the pair that they needed to restart that maneuver and make another attempt. During analysis, number of errors was determined by the number of times the participant pair had to restart a maneuver within a trial.

I expected to observe a decrease in the time it takes to complete each trial as the participants went through the trials. Unfortunately, I was unable to conduct this analysis. Time on trial was influenced by technical errors such as recharging the remote-control car and by different ability levels by participants to drive the remote-control car. Time on trial was also influenced by conversations with the experimenter that were removed for analysis purposes. As I am unable to control for these factors, I did not conduct this analysis and do not make any conclusions regarding time on task.

Procedure

First, the participants entered the experiment room together. The participants were assigned a role based on a coin toss. [3]The participant assigned the role of driver had 15 minutes to familiarize themselves with the controls of the car. Next, the director watched a video of 6 different clips of maneuvers executed by the car. After watching the clips, the director instructed the driver on how to correctly execute the string of 6 actions. Once the pair successfully completed the first trial, the driver and the director repeated the same procedure five more times. Each new trial had the same action clips, but in different orders. The first and last trial were set as described in the task section and trials two through five were randomized for each participant pair. The conversations were audio recorded and the actions of the participants were video recorded. The audio recordings were transcribed for the analysis.

The experiment concluded once the participants completed all six trials or when the allotted two hours elapsed, whichever came first. The pairs that did not finish in the two hours were not included in the data analysis. At the end of the experiment or the two hours, participants were thanked for their participation, given an explanation of the purpose of the study, and provided an opportunity to ask any follow up questions.

Design

This study is a repeated measures design with 6 maneuvers per trial, and 6 trials. The 6 maneuvers were broken down further into type of maneuver: path, manner, and a

[3] I did not lead directed training on the cars because that could potentially introduce bias towards specific words during the experiment.

combination of path and manner. Two of each maneuver type created the list of 6 maneuvers. In total, the experiment took each participant pair one to two hours to complete.

Preliminary analysis

Data cleaning. Four pairs that failed to complete all six trials were removed from the data set. Two pairs were removed due to not meeting the English language requirements and failing to provide TOEFL scores as requested. Eight pairs were removed because the video and audio recording device did not fully record the data – four were battery issues, two were storage issues, and two were audio issues. Then the transcriptions from the participant pairs who completed all six trials were cleaned to ensure only task relevant conversations were analyzed. I removed any dialogue exchanged between the experimenter and the participants. For example, one participant asked the experimenter how long the study will last. The question in addition to the experimenter's response were removed from the analysis. All dialogue from the experimenter was removed from analysis. The amount of dialogue removed varied by participant pair, but on average 10%-15% was removed.

After the data was cleaned, I used the LIWC program (Pennebaker et al. 2015) to analyze the linguistic characteristics of the dialogue between the director and the driver in this experiment.

Results

General Trial Effects

Consistent with previous work, I anticipated overall declines in word counts across the trials as well as overall declines in verb phrase count. I also tested for an anticipated positive trend in accuracy and an anticipated decline in the number of turns (alternations in speech) taken by each participant in the pair. As an extension on previous work, I tested linear, quadratic and logarithmic models and have reported significant trends below. For all tested models, see appendices B through M.

Verb phrase count. As predicted in Hypothesis 1, I observed a linear decline in verb phrase count as participants repeated the six-maneuver sequence for six trials. To test this, I analyzed the proportion of verbs within each trial and modeled the decline. Logarithmic and quadratic models were also tested but were not significantly different from the linear model. Multiple regression analysis indicated that together, trial and subject pair explained 19% of the variance (*Adjusted* R_2 = .19, $F(2, 63) = 8.48$, $p <$ .001). See Table 1 for a summary of the results.

Table 1. *Regression Analyses of Participant Pair and Trial on Verb Phrase Count*

Predictors	β	*Adj.* R_2
Step 1		.19***
Subject	-.31**	
Time	.34**	

Note. *$p < .05$, **$p < .01$, ***$p<.001$

The example below illustrates the difference between the verb phrase (phrases highlighted) usage from trial one to trial six (Combination maneuver – Pair Videos 0058, 0059):

> Trial 1
> Driver: Ok
> Director: On the x, and then you're going to make, you're going to go...right not completely, not straight through, toward the red thing a little bit, and then go between the cone by the basketball and the one on the edge and then toward the mirror, don't touch the mirror, and go around the cone between the basketball and the x
> Driver: Around the cone but between the basketball and the cone?
> Director: Between, ok, there is the basketball and there is the cone go around the cone two times, and then you're going to stop when the car is in front of the cone, closest to you…
> (Continuing conversation with less verb density)
>
> Trial 6
> Director: And then go around the cones two times and then stop in the front
> (Successfully completed maneuver)

Word count. As predicted in Hypothesis 2, I observed a logarithmic decline in total word count as participants repeated the action sequence for each trial. A logarithmic regression model identified a statistically significant reduction in the mean number of total words used by participants in each trial (*Adjusted* R^2 = .65, $F(2, 63) = 41.64$, $p < .001$). The model is as follows: y = (-.06)Participant Pair + (1.41)Trial + (-2.10)log(Trial) (See Figure 2). When fit with a linear model excluding the first trial, less of the variance is accounted for (*Adjusted* R^2 = .27, $F(2, 52) = 11.18$, $p < .001$). When fit with a logarithmic model excluding the first trial, more variance is accounted for (*Adjusted* R^2 = .32, $F(3,51) = 9.32$, $p < .001$) than in the linear model, but less than in the

logarithmic model including the first trial. Therefore, the logarithmic model is the model of best fit.

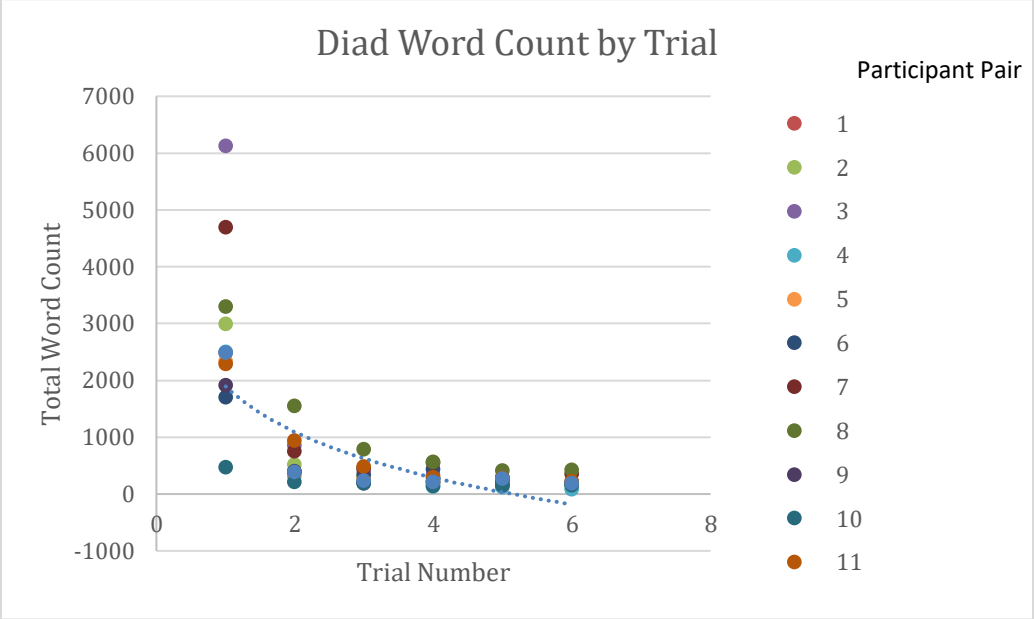

Figure 2. Word count per trial per participant pair

Task accuracy. I observed a logarithmic decline in the number of errors as the participants progressed through the trials (*Adjusted* R^2 = .71, $p < .001$). This decline supports Hypothesis 3. Task accuracy was measured by number of errors committed by the participant pairs. Number of errors was defined as the number of incorrect attempts made by the pairs. Each pair was required to successfully complete each maneuver to be able to move on to the next maneuver. The figure below (Figure 3) illustrates the number of errors by participant pairs for each trial. Errors declined most from trial one to trial two, and no pairs had more than two errors from trial two onwards. When fit with a linear model excluding the first trial, little variance is accounted for (*Adjusted* R^2 = -0.01,

$F(2, 52) = 0.78$, $p > .05$). When fit with a logarithmic model excluding the first trial, less variance is accounted for (*Adjusted* $R^2 = 0.14$, $F(3, 51) = 3.92$, $p < .05$) than the logarithmic model including the first trial, but more variance than the linear model without the first trial. Therefore, the logarithmic model including the first trial is the model of best fit.

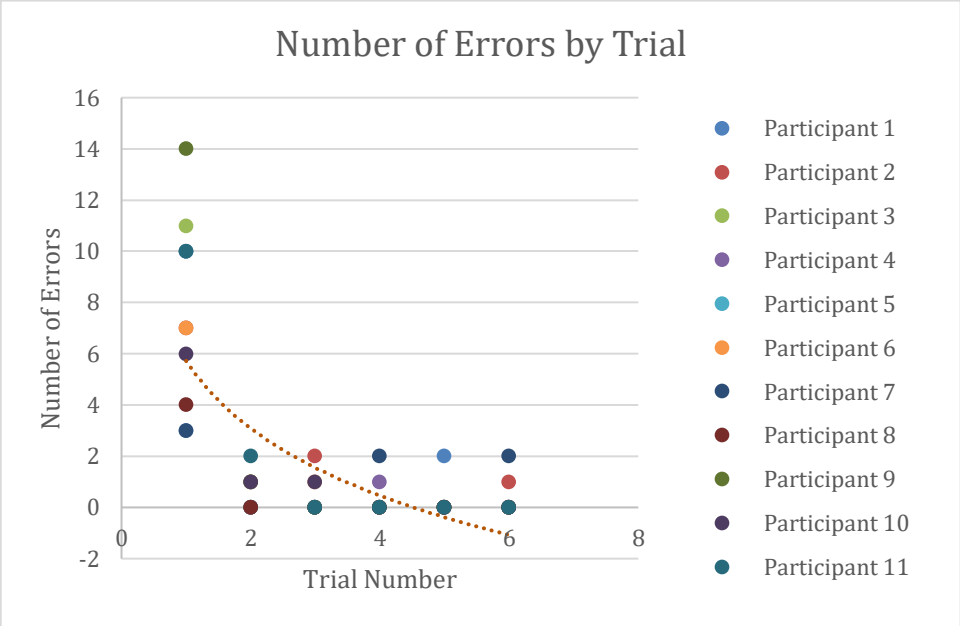

Figure 3. Number of errors per trial per participant pair

Turn-taking. The measure I used to identify mutually agreed upon terms is the number of times the participants alternated speaking, also referred to as turn-taking. As the participant pairs progressed through the trials, the number of turns taken decreased logarithmically (*Adjusted* $R^2 = .58$, $p < .001$). This decline provides support for Hypothesis 4. In Figure 4, the dramatic decline from trial 1 to trial 2 is demonstrated.

When fit with a linear model excluding the first trial, less variance is accounted for (*Adjusted* R^2 = .14, $F(2, 52) = 5.34$, $p < .01$). When fit with a logarithmic model excluding the first trial, less variance is accounted for (*Adjusted* R^2 = -0.02, $F(3, 51) = 0.69$, $p > .05$) than in the linear model without the first trial and in the logarithmic model including the first trial. Therefore, the logarithmic model including the first trial is the model of best fit.

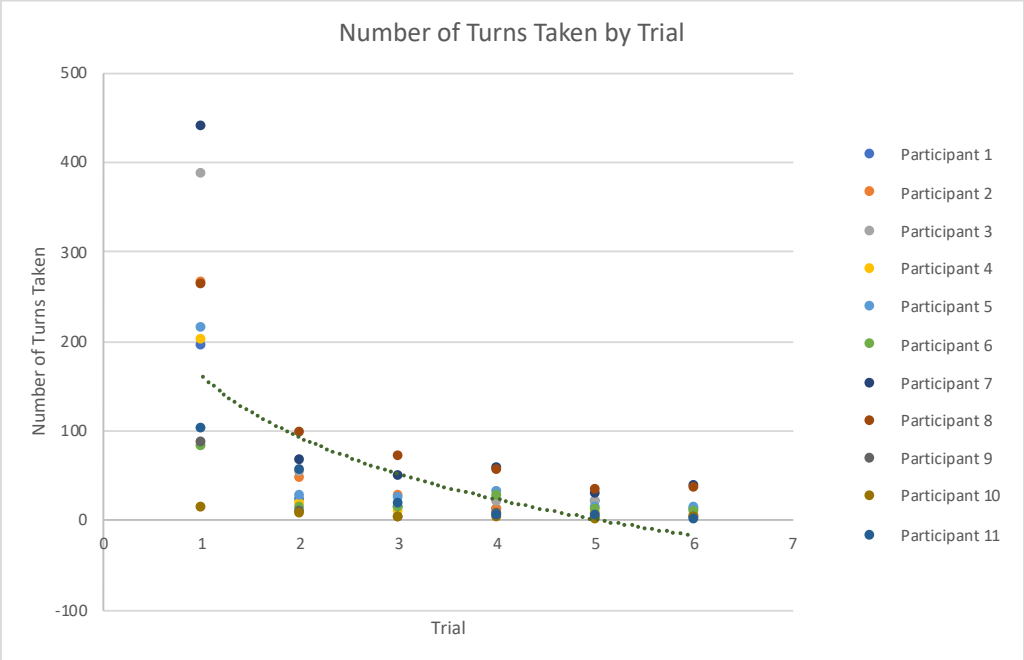

Figure 4. Number of speech turns taken per trial per participant pair

Below is an example of the dialogue between a director and driver for the same maneuver in trial one and then again in trial six. The example demonstrates that the number of turns taken decreased from trial one to trial six, but also shows the phrase the

two participants had come to understand to initiate that maneuver. In trial one, you can see the words "ever so slightly" and "bottom of the mirror" that appear again in trial six.

Trial 1:

Director: You're gonna start driving towards the mirror ever so slightly go to the left and then straighten back out, so you go right in the middle between the basketball and the closest cone to its right
Driver: Ok
Director: And then just go all the way forward until the wheels touch the bottom of the mirror, and then stop there
(Driver successfully completed maneuver)

Trial 6:
Director: The first maneuver is the same one we just did where you go up to the bottom of the mirror ever so slightly
Driver: Not too hard
(Driver successfully completed maneuver)

Maneuver Type Effects

To expand upon Clark and Wilkes-Gibbs' research, I sought to analyze the dialogue by breaking down the conversations for specific types of maneuvers. The categories hypothesized to have an impact on action tasks consisted of prepositions, auxiliary verbs, adverbs, compare, insight, discrepancy, verbs, tentative, certain, focus on past, focus on present, focus on future, motion, space, time, assent, and interrogatives. Verb count and total word count were used in analyses of trial effects by maneuver type. Each pair's dialogue was analyzed separately by each trial. LIWC provided output of the proportion of each category within each trial to ensure independence from the decline in total word count. Table 2 illustrates which linguistic measures had significant linear or

nonlinear trends in the whole dialogue as well as by maneuver type (see Appendices B - M for all maneuver type effects).

Table 2. Summary table of model results examining trial effect by maneuver type.

Measures	Manner	Path	Combination	Whole Dialogue
Word Count	-Nonlinear	-Nonlinear	-Nonlinear	-Nonlinear
Focus on Past	-Linear	-Linear	-Linear	-Linear
Verbs	NS	-Linear	-Linear	-Linear
Discrepancies	NS	-Linear	-Linear	+Linear
Tentativeness	NS	-Nonlinear	NS	-Nonlinear
Prepositions	NS	-Linear	NS	NS
Focus on Future	NS	-Linear	NS	NS
Certainty	NS	NS	-Linear	-Linear
Causalities	-Linear	-Linear	NS	+Linear
Focus on Present	NS	NS	NS	+ Linear
Motion	NS	NS	NS	NS
Space	NS	NS	NS	NS
Time	NS	NS	NS	NS
Assent	NS	NS	NS	NS
Interrogatives	NS	NS	NS	NS
Auxiliary Verbs	NS	NS	NS	NS
Adverbs	NS	NS	NS	NS
Comparatives	NS	NS	NS	NS
Insight	NS	NS	NS	NS

Note: NS = Not significant, +/- = Direction, linear/non-linear = Type of relationship

Consistent characteristics. When each maneuver type was analyzed separately as well as all together as one, two linguistic categories were consistent across all analyses: word count and the proportion of past-focused words. Word count had a nonlinear relationship as participants progressed through the trials. The proportion of past-focused words used had a negative relationship as the participants progressed through the trials.

Path maneuver effects. The most prevalent pattern in results of maneuver type effects is within path maneuvers. Seven of the ten linguistic categories that had significant relationships reflected path effects. Proportion of prepositions, verbs, discrepancy related words, and future-focused words all had negative relationships in path maneuvers. Specifically, in proportion of prepositions and of future-focused words, only path maneuvers displayed this negative linear relationship compared with manner, combination, and even the dialogue as a whole. For example, in the path maneuver dialogue below, the director and the driver use prepositions in their questions to each other to achieve common ground. The path maneuver example below illustrates anecdotally that the proportion of prepositions used in trial one (16.13%) decreases in trial six (9.47%).

> Path
> Trial 1
> Director: Ok, so bear with me on this one, it's kind of challenging. So, um, you'll go. Let me start over. So it should be two cones next to each other but by the basketball
> Driver: Ok
> Director: That's like in *the L part*

Driver: Mhmm
Director: So you'll go in between *those two cones* by *the basketball*
Driver: Ok
Director: Are you there?
Driver: Not yet, go between the cones?
Director: Yea
Driver: Ok
Director: Um, I think this is the left, um, you go around the cone that has the basketball next to it, you'll make a circle
Driver: Around the basketball or just the cone?
Director: The cone that's next to it.
Driver: Ok
(Dialogue continues with less preposition density)

Trial 6
Driver: Give me a second. Ok
Director: So for the next one you're gonna go in between *the two cones of the L shape* you're gonna go make a full circle and then you're gonna make a half circle and then you'll stop, you should be in front where the X is
Driver: Ok
(Driver successfully completed maneuver)

The manner maneuver example below illustrates a different picture than the path maneuver. The proportion of prepositions used in trial one (9.38%) is not significantly different from the proportion of prepositions used in trial six (10.77%).

Manner
Trial 1
Director: Oh this one should be, oh this one should not be bad. Alright, you ready?
Driver: Yea
Director: Uh, so start on the x facing the mirror
Driver: Yes
Director: You're gonna start driving towards the mirror ever so slightly go to the left and then straighten back out, so you go right in the middle between the basketball and the closest cone to its right
Driver: Ok
Director: And then just go all the way forward until the wheels touch the bottom of the mirror, and then stop there

(Dialogue continues with less preposition density)

Trial 6
Driver: Yes
Director: The first maneuver is the same one we just did where you go up to the bottom of the mirror ever so slightly
Driver: Not too hard
(Driver successfully completed maneuver)

It appears that path maneuvers require more prepositions (and their arguments) at the outset to describe than the manner maneuvers.

As mentioned in the trial effects on proportion of verbs previously, there is a negative linear relationship between proportion of verbs and trial number. When the maneuver types are analyzed individually, dialogue from path and combination maneuvers are the source of the verb decline. Dialogue from the path maneuvers are also the source of the negative linear relationship in the proportion of future-focused words and trial number, and the negative nonlinear relationship between the proportion of tentativeness words and trial number. The dialogue as a whole did not exhibit a significant change with future-focused words, but a negative linear relationship does emerge when path maneuvers are analyzed separately. Manner and combination maneuvers did not show the same relationship.

Interestingly, path and combination maneuvers display negative linear relationships between trial and proportion of discrepancies. But when the dialogue is analyzed as a whole, the relationship changes direction to positive. This suggests that even though the driver and director achieve some common ground during the repeated

tasks, there are still some discrepancies within the dialogue. Below is an example of a path maneuver:

> Trial 1
> Director: Alright you want to drive toward the corner of the red ramp and then go in front of the basketball until like a little circle but staying inside the cone towards the white circle
> Driver: Is that do I just go around? Can you repeat that?
> Director: Yea, um, you want to go towards the corner of the ramp and then you use that make a go in front of the basketball and it's you're kinda going in like you're following the cones in like a little a little circular pattern toward the white circle so you can put the car partly on the white circle.
> (Dialogue continues with limited use of discrepancy related words)
>
> Trial 3
> Director: Ok, you want to go towards the edge of the ramp near the basketball and then go and then circle a circle on the inner part of the cones to the right to the white square, circle
> (Successfully completed the maneuver)

Relatedly, manner and path maneuvers separately display negative linear relationships between causality-related words and trial number, but a positive linear relationship in the whole dialogue. The example below shows the director using one causal phrase in trial one and then using a similar causal phrase while overall, the director uses less words to convey his/her directions thus increasing the proportion of causality-related words as the participants progressed through the trials.

> Trial 1
> Director: Ok, so for this first maneuver, you're gonna be on the X facing the mirror and it's gonna be basically a looping u turn, you're gonna start off looping to the left, make sure you turn *before the basketball* and bring it all the way back down *to the white circle*
> Driver: Ok, so going
> Director: So loop to the left, come all the way around, stay inside the cones and then just go straight for the white circle

Driver: Oh just stay inside the cones?
Director: Um, it's kind of the car's front right tire in the white circle…
(Continuing conversation with less causality related word density)

Trial 3
Director: Alright, as we get closer to the X, can I tell you how to do this next one?
Driver: Ok
Director: It's the looping to the left you make the big u turn and end with the front right tire *on the white circle*
Driver: Ok, um
(Successfully completed maneuver)

Manner maneuver effects. The only measures that were affected in manner related maneuvers were proportion of causality-related words, word count, and proportion of past-focus words. Word count and past-focused words were affected by all three types, but proportion of causalities was the only linguistic category that was affected by manner specific maneuvers as well as path maneuvers, that was not affected by combination maneuvers. This suggests that causal words are integral to successfully communicating action or that causal words hold more information than other words that decline or disappear. Below is an excerpt from a pair completing a manner maneuver:

Trial 1:
(Earlier exchanges removed)
Director: Oh gosh, it's so it's so repetitive, I don't know how many times, hold one. Ok so you *went forward and now you went backward* now you're gonna go forward again past the red X, stop at the same exact spot, so front tires in front of the red, front tires
Driver: At the red or in front?
Director: At the red, so *not in front, front tires* does that makes sense? You're gonna you're gonna keep going back and forth across the red X, the X, I don't know why I keep calling it a red X and then you're gonna go back, so you stopped at the front and you're gonna go back past the X, now I have to go back, how

many times have we done that? We went forward, you went backward, we went forward, we went backward, ok you're gonna go forward again third time back tires are gonna go align with the not the front tires, the back tires this time
Driver: Alright, align with the front of the red tent?
Director: Yea, and then past the red back past the X, so this time you're just *going forward and backward* no left and right, make sense?
Driver: Alright
(Further exchanges removed)

Trial 6
Director: Ok so the next one is forward and backward
Driver: Alright, front, back
Director: Front again
Driver: Back, back?
Director: Yea, back back, have you done back twice?
Driver: Yea
Director: And then um, then it's front
Driver: Uh huh
Director: Back and then you go barely and then end on the X
(Pair successfully completes maneuver, no causality related words used)

Combination only maneuver effects. The proportion of certainty-related words in the combination type of maneuver has a negative linear relationship with trial. The example below shows that as common ground is formed, certainty words are not necessary after truncation of the phrases. Example below:

Trial 1
Director: Ok, make sure the car so, um, there is a left path and a right half from the X
Driver: Mhmm
Director: When you're looking at it. Make sure that the body of the car is on the left half of the X
[dialogue that did not contain certainty words removed]
Driver: Ok
Director: And stop at like that cone um over there. Now um, and make sure it's kind of in the middle between ok, we might have to restart this one, but now you

know, you know how like how it's like there's a space in between that cone and that cone and that *um red piece?*
Driver: Mhmm
Director: It's like *a little triangle type thing*, you want to make sure the car is in the middle of that. Um and then and then turn to the left um left turn in towards that um like turn so the cars vantage point um turn left into that um like you're like you're about to go into that um cardboard but then as soon as you um get there, make sure your car is straight again and it should be coming towards you like the uh yea. And you just finish at there

Trial 6
Director: Ok, alright start on the X on this one, go towards that um *red tent looking thing*, the corner and go straight back along the line and finish at the end there
(Successfully completed maneuver, no certainty words)

Present-focus related words also had an interesting relationship. When analyzed separately, none of the three maneuver types had a significant relationship with trial number, but the overall dialogue displayed a positive linear relationship.

Discussion

The purpose of this study was not only to replicate the original results of Clark and Wilkes-Gibbs (1986) using objects to action tasks, but to expand upon the results by identifying the specific linguistic properties of the dialogue that influence the creation of mutually agreed upon terms/phrases for action. I also expanded upon Clark and Wilkes-Gibbs (1986) original study by analyzing and comparing logarithmic, quadratic, and linear models to describe the relationships between variables. Finally, I looked further into the classification of the different maneuvers by type of action (manner related, path related, and a combination of manner and path) and analyzed them separately to identify differences in patterns.

In general, my study using action is consistent with the original study using objects. I observed significant declines in total word count, errors, and turn taking between the participants in the pair with number of trials. Verb phrase count also declined with a similar pattern to the noun phrases in the original study. These findings were still replicated even with the expanded focus on action.

Analyzing linguistic characteristics of the three types of maneuvers shed light on the language differences that exist when explaining path related maneuvers versus manner related maneuvers. Path maneuvers required larger proportions of prepositions in earlier trials, but then decreased in later trials. While I did not make any official a priori hypotheses relating to the linguistic characteristics, the decrease in prepositions supports a linguistic (and potential conceptual) distinction between Schank's PROPEL (Manner in this study) and PTRANS (Path in this study) conceptual primitives (1972). I believe that this is due to the increased need to explain actions in relation to objects in the environment (before the ball, next to the tent) in early trials to establish common ground relating to the environment, affecting metacognitive measures.

Overall trial effects

Replicated effects. The replication predictions in Hypotheses 1, 2, and 4 were overall supported. The linear decline in verb phrase count over the six trials is consistent with Clark and Wilkes-Gibbs' finding of noun phrase decline (1986). The similar declines in both experiments reinforce the linguistic theory that noun and verb phrases are governed by the same underlying structure (Chomsky, 1970). The logarithmic

decline in total word count is also consistent with Clark and Wilkes-Gibbs' findings (1986). The steepest decline from trial one to trial two and almost no decline from trial five to trial six mimics the findings and suggests that the pattern shown in Clark and Wilkes-Gibbs (1986) can be extended into action-based tasks. The decline in speaking turns for each participant suggests the driver and director communicated more in the first trials and then the need to communicate declined when the pair had mutually agreed upon certain words and/or phrases. In the last trial, the number of turns taken was often zero. This is also consistent with the original study.

Task accuracy. Task accuracy across trials was not analyzed in Clark and Wilkes-Gibbs' (1986) study, but the overall reported error rate was 2%. I included task accuracy in my hypotheses to address the added complexity introduced by action tasks. The results provide support for Hypothesis 3. The added complexity explains the increased error rate in this study compared to the 2% in the original study. I found that the number of errors significantly decreased as the participants progressed through the trials. Participants had the highest error decline from trial one to trial two, and no pairs had more than two errors from trial two onwards.

Common ground. The significant reductions in word count, verb phrase count, and errors is consistent with the original explanation that participants are forming common ground with their partners to achieve their shared goal of completing this action-based task. Clark and Wilkes-Gibbs (1986) identified common ground as one of the

main explanations of their findings and I believe this study can also point to common ground as the explanation for the results.

Maneuver Type Effects

While I confirm the general pattern of results with verbs that Clark and Wilkes-Gibbs (1986) identified with nouns, a closer look at the sessions by maneuver type helps to identify a possible dependence of this finding on the nature of the entity being described. In doing so, I used more refined measures of language type than Clark and Wilkes-Gibbs (1986) employed. The single most compelling and coherent result emerging from this study is that language reduction is associated with path maneuvers.

Trends in word count and past-focused words were consistent across all three maneuver types. Looking at the summary table in the results section, overall path and combination maneuvers have significant relationships in more linguistic categories than manner maneuvers when considered individually. This suggests that path components of action tasks are responsible for reduction in words used, and therefore leads to truncation into mutually agreed upon phrases.

The negative nonlinear relationship of tentative words in the whole dialogue was influenced by path maneuvers rather than manner and combination maneuvers. I theorize that this is due to how we describe path-related action compared to manner. In the first trials, participants need to explain movement in relation to objects in the experimental environment. This introduces a greater use of tentativeness related words

while the driver and director are exploring the environment and establishing common ground.

The decrease in future focused words in path maneuvers suggests that participants initially needed to describe where the car should go next, but then were able to drop those instructions in the truncation process. Interestingly, present-focused dialogue increased linearly when the whole dialogue was analyzed together, but each of the types alone had non-significant relationships. Future research should explore this relationship to determine the underlying reason behind this.

The proportion of prepositions was only significant in path maneuvers, where there was a negative linear relationship. Explaining path maneuvers requires more prepositions to describe where the car moves in relation to objects in the experimental environment. Prepositions such as next, between, by, and around were used to establish exactly where the car was traveling and enabled the driver to successfully complete the maneuver. The decrease suggests as the participants moved from trial to trial, they were able to truncate longer directions that included a lot of prepositional phrases into short, concise phrases that only preserved the most important prepositions. Even though the English language relies on prepositional phrases when discussing path-type maneuvers, the amount necessary to direct a partner with shared common ground requires significantly less than necessary to direct without that shared common ground. Manner and combination maneuvers did not have significant relationships, but this does not indicate that participants did not use prepositions. Rather, it suggests that the

prepositions used in trial one were preserved to trial six and were not dropped in the truncation process.

Proportion of certainty words had a significant negative linear relationship when dialogue was analyzed as a whole, and when analyzed separately, in combination maneuvers. This suggests that in more complex maneuvers the proportion of certainty words are affected by trial progression while the simpler maneuvers are not.

The proportion of discrepancy, causality, and present-focused dialogue do not follow a clear pattern. In all three linguistic categories there are conflicts between what is found when the dialogue is analyzed as a whole compared to when the maneuver types are analyzed separately. One potential source of this conflict could be that due to the proportional nature of the results, later trials result in an increase in proportion of discrepancies and causalities when manner maneuvers are included. Discrepancies and causalities might carry more information in manner maneuvers. The discrepancy measure could be measuring error at the outset, but then also goals at the end. Further research is needed to investigate if these findings are due to an underlying phenomenon or if they are spurious results.

The other categories measured and analyzed (motion, space, time, assent, interrogatives, auxiliary verbs, adverbs, comparatives, and insight) were not significantly affected by trial or by individual maneuver types. Lack of significant effects can either imply that words associated with those categories were important and could not be

condensed or that words in those categories are not used or dropped in any systematic pattern. Further research can provide insight.

Overall, the patterns that were observed suggest that aspects of path maneuvers significantly affect the linguistic characteristics of the dialogue used to convey the action. During the process of achieving common ground and truncating the directions, individuals are able to eliminate excess directions in path maneuvers and still achieve the same successful outcome of finishing the maneuver. On the other hand, manner maneuvers inhibit truncation. There is evidence in the dialogue that suggests that this is because in general, manner maneuvers do not require as much referential language that can be dropped off in later trials. Moreover, learning the structure of the environment is less important. This is consistent with the need for location specific language with PTRANS action primitives and not as necessary for PROPEL action primitives (Schank, 1972).

Contributions

The purpose of my study was to analyze the development of language for action, with implications for distributed planning and to determine patterns of language development involved in the process. This study contributes to the understanding of compositional language for novel action. Currently, there are no studies focusing on action sequences, and this study aimed to address this. My study found that participants decrease in the total words used, the number of verb phrases, and the number of errors committed. These results raise issues relating to the formation of new phrases between

people attempting to reach a common goal and the efficiency of these practices. Understanding the process of creating common ground offers insight into how we develop schemas collaboratively when working towards a common goal as well as provides patterns to take into consideration when developing planning documentation tools for dynamic action-based tasks.

My study also identified that patterns in linguistic categories differ between different types of maneuvers. From a practical perspective, my findings aim to increase efficiency of communication in distributed planning either between humans, or humans and machines. From both a theoretical and practical perspective, different types of action (manner vs. path) require different language to explain and eventually achieve common ground. Repeated path tasks in the same environment will eventually result in abbreviated language that follows a visible pattern in linguistic characteristics. While repeated manner tasks do result in a decrease in overall words, there aren't any specific language characteristics that were only affected by manner maneuvers, at least for the number of repetitions studied here.

Issues and Limitations

I controlled for the potential influence of the laboratory setting by mimicking a real-world task---the operation of the Mars Rovers. I did not however, control for partner familiarity, which could also be important. Participant knowledge of remote-control car functioning is another limitation that may have influenced my results. I did provide the driver five minutes to become familiar with the controls and understand how the car

moves through the space. However, differences in familiarity with remote controlled cars may have added noise that overwhelmed important trends. I did not conduct a formal training session with the participants to eliminate the risk of influencing the participants' language. I suggest future researchers seek out participants who have experience with remote control cars. This will eliminate the learning curve at the beginning of the experiment and may help identify stronger patterns.

The remote-control car itself constituted another limitation. It has a battery life of about 45 minutes, which is quite long in terms of remote-control cars, but insufficient for a two-hour experiment. I used two identical cars to eliminate wait time for the remote-control car to recharge. This did not completely eliminate breaks between trials. During such breaks, participants were instructed not to speak to each other, especially about the task. Two additional cars did not eliminate the issue either, as two of the cars became completely nonfunctional and had to be replaced. Future research should consider using a simulation task or other action-based tasks that do not rely on battery operated equipment. Future researchers can also design the experiment around the limitations of the remote-controlled car.

This study included unforeseen variables such as stopping and starting of the vehicle, experimenter interventions, and extra dialogue between the participants and the experimenter. The task was also significantly longer than the original Clark and Wilkes-Gibbs (1986). Because each maneuver took longer to execute than identifying an image and placing it in the correct place in a line up, the participants had extra communication

time even if the driver was able to identify the maneuver with shorter phrasing. I believe this may have been the cause of longer phrases persisting in subsequent trials rather than specific names for the maneuvers seen in Clark and Wilkes-Gibbs (1986). This fact combined with a similar pattern of decrease in turn taking suggests common ground was still achieved in the longer phrases.

Unfortunately, Clark and Wilkes-Gibbs (1986) did not provide readers with much explanation on what specific analyses they used to identify the relationships, and I could not determine the error terms. Thus, the analyses I conducted may not be direct replications of the original study, but the general patterns are replicated.

While these results mimic what Clark and Wilkes-Gibbs found in 1986, results from both studies can not be blindly applied to other stimuli or domains without acknowledging the fact that these results were found in reference to these specific tasks, stimuli and crucially, an experimental environment and may not necessarily be consistent in other domains or with other stimuli. The language might be a function of the set of stimuli and not the universal set. This is as relevant to Clark and Wilkes-Gibbs as it is to my study. Therefore, all the models are fixed effects models with subject selection being a random effect.

Future research

Future research should explore action language across different domains and within domains between different environments. Conducting similar studies in different environments will help determine if the relationships that were found in the language

used are a function of specific environmental cues of this experiment or if they are in fact influenced by the properties of path maneuvers. Learning the environment could lead to the truncation of phrases, so researchers should investigate this by having subjects repeat the same maneuvers in different environments.

The real-world example that inspired this research was telerobotics in space, specifically the Mars Rover Expeditions. The lack of focus on the development of accepted methods within the planning documentation on that mission introduced challenges later when attempting to identify and recover action sequences of the Mars Rovers that had evolved over time. Team members assigned accepted methods to repetitive motion sequences over time as they gained experience with the maneuvers. Interestingly, in my study I occasionally observed participants agreeing upon terms and phrases that had little to no relevance to the specific motions within the maneuver. Instead, some pairs referred to maneuvers as “favorite one”, “easy one”, or “the really complicated one”. This anecdotal evidence further supports the need for a focus on the development of accepted methods over time because it would be impossible to reconstruct a maneuver with such little information about the specific aspects of the maneuver.

There are many applied problems and domains apart from space exploration that can benefit from this research. One of the more pervasive examples is the acquisition of surgical skills. Recent research aimed to identify ways to increase the effectiveness of remote instruction in surgery (Mackenzie et. al., 2015; Shah et. al., in press). Motions in

surgery consist of both path and manner related movements. Surgeons and remote instructors not only use patient specific anatomical landmarks to guide these movements, but they also instruct how much pressure to apply in certain motions, how to hold instruments, and how the muscles should feel. Patient-specific anatomy also brings in the issue of generalization from one patient (environment) to another whose anatomy might differ drastically. My results can be used to inform researchers of the important differences between language necessary to instruct path and manner related surgical movements and create effective and efficient training processes.

Another aspect for future research to investigate is the issue of retention. While my results suggest that common ground is formed in the six trials between the participants, I do not know if the truncated phrases persist over time. The phrases could be a function of short-term convenience, or the common ground could persist or extend to other similar tasks. I suspect that there is a period of retention immediately following the initial task, but the longer the time between repetitions, the more common ground needs to be reestablished. Future research should also attempt to address the issues and limitations in the previous section where I have laid out recommendations to do so.

Conclusions

Theoretical. The purpose of my study was to explore the different patterns and relationships in the planning process of action. I found results in these specific action-based tasks that are consistent with Clark and Wilkes-Gibbs' (1986) study on object-based tasks. Total word count and verb phrase usage significantly decreased as

participants progressed through the experiment. The number of turns each participant took to speak also decreased significantly over time. This suggests that like in Clark and Wilkes-Gibbs, participants communicated more effectively over time after they achieved common ground and were able to create new agreed upon terms to describe the different action-based tasks in the experiment. My results support a general distinction between path and manner verbs by showing different associated language patterns and explain the learning of path maneuvers as dependent on learning the features in the environment in which they are executed.

Practical. My results assist with interface design in action-based tasks. The natural progression of language usage between collaborators informs designers to take into account the flexibility of describing and naming tasks and create flexible interfaces that can adapt to describing and naming changes. My results are also useful for improving planning for major action-driven tasks in various workplaces. Understanding that in the beginning phases of action-based tasks, there is a necessity for more time and more space in documentation to accommodate achieving common ground. Even though there were some issues and limitations that should be addressed in future work, I believe this study begins to fill the action-based task gaps in the common ground literature and will be used to inform planners and designers in their work practices.

Appendix A: Demographics Questionnaire

Age: _____

Gender: _____

Race: _____

Approximate number of college credits completed: _____

Native Language: _____

If English is not your native language, please provide your TOEFL scores for the following:

Listening: _____ Speaking: _____

Major: _____

Do you play video games? (circle one) Yes No

Hobbies: _____

Do you have experience with remote controlled vehicles? (circle one) Yes No

Work experience: _____

Appendix B. *Multiple Regression – Whole Dialogue*

Measures	*M*	β_{Trial}	β_{Pair}	*F*	*Adj. R*2
Word Count	744.91	-.62***	-.06	19.85***	.37***
Tentativeness	.79	-.28*	-.10	3.09	.06
Focus on Present	11.05	.27*	.34**	7.38**	.16**
Prepositions	13.53	-.18	-.54***	15.27***	.31***
Causalities	.46	.39**	.20	7.50**	.17**
Discrepancies	.63	.30*	.30**	7.021**	.16**
Verbs	14.71	-.31**	.34**	8.48***	.19***
Certainty	.54	-.36**	-.19	6.32	.14**
Focus on Past	1.32	-.36**	-.20	6.44**	.14**
Focus on Future	3.76	-.09	.27*	2.82	.05
Motion	7.33	.10	.45***	13.84***	.28***
Space	15.16	-.03	.09	.26	-.02
Time	10.51	.27*	-.13	3.10	.06
Assent	4.80	.02	-.16	.87	-.00
Interrogatives	1.43	-.01	-.53***	12.00***	.25***
Auxiliary Verbs	6.69	-.04	-.21	1.54	.02
Adverbs	7.71	.04	-.14	.65	-.01
Comparatives	1.95	-.25*	-.08	2.38	.04
Insight	.51	-.16	-.10	1.85	.01

Note. *$p < .05$, **$p < .01$, ***$p<.001$

Appendix C. *Multiple Regression – Path Maneuvers*

Measures	*M*	β_{Trial}	β_{Pair}	*F*	*Adj.* R_2
Word Count	235.47	-.57***	-.00	15.49***	.31***
Tentativeness	.71	-.36**	-.02	4.88*	.11*
Focus on Present	12.47	-.03	.19	1.25	.01
Prepositions	13.26	-.24*	.42***	9.89**	.21**
Causalities	.69	-.17*	.04	3.66*	.08*
Discrepancies	.46	-.38**	.21	7.24**	.16**
Verbs	16.05	-.19	.25*	3.58*	.07*
Certainty	.49	-.12	-.02	.47	-.02
Focus on Past	1.20	-.30*	-.21	5.01**	.11**
Focus on Future	3.28	-.10	.31*	3.73*	.08*
Motion	7.38	.16	.29*	3.83*	.08
Space	13.29	-.10	.37	2.24	.04
Time	8.51	.32**	.02	3.70*	.08*
Assent	5.21	.05	-.18	1.14	.00
Interrogatives	1.54	-.02	-.16**	5.27	.12**
Auxiliary Verbs	7.18	-.09	-.09	.54	-.01
Adverbs	7.18	.07	-.02	.18	-.03
Comparatives	1.66	-.18	.04	1.11	.00
Insight	.51	-.14	-.06	.73	-.01

Note. *$p < .05$, **$p < .01$, ***$p<.001$

Appendix D. *Multiple Regression – Manner Maneuvers*

Measures	*M*	β_{Trial}	β_{Pair}	*F*	*Adj.* R_2
Word Count	215.23	-.49***	.04	9.82***	.21***
Tentativeness	.87	-.04	-.03	.09	-.03
Focus on Present	10.23	-.20	.34**	5.80**	.13**
Prepositions	11.01	-.13	.48**	6.45**	.14**
Causalities	.36	-.25*	.05	2.19	.04
Discrepancies	.79	-.13	.31*	3.98*	.08*
Verbs	14.05	-.21	.34**	6.09**	.14**
Certainty	.59	.05	-.11	.48	-.02
Focus on Past	1.37	-.26*	-.04	2.38	.04
Focus on Future	4.17	-.04	.09	.31	-.02
Motion	8.33	-.03	.48***	9.44***	.21***
Space	14.75	-.07	.18	1.17	.01
Time	13.40	.11	-.09	.68	-.01
Assent	5.38	.13	.14	1.20	.01
Interrogatives	1.46	.03	-.40**	5.92**	.13**
Auxiliary Verbs	6.99	-.11	-.05	.48	-.02
Adverbs	7.62	-.09	-.25	2.45	.04
Comparatives	2.28	-.24	-.20	3.38	.07
Insight	.67	-.03	-.03	.06	-.03

Note. *$p < .05$, **$p < .01$, ***$p<.001$

Appendix E. *Multiple Regression – Path-Manner Combination Maneuvers*

Measures	*M*	β_{Trial}	β_{Pair}	*F*	*Adj.* R_2
Word Count	295.09	-.55***	-.13	14.73***	.30***
Tentativeness	.70	-.28*	-.05	2.74	.05
Focus on Present	10.90	-.23	.24	3.82*	.08*
Prepositions	15.70	.11	.40***	6.68**	.15**
Causalities	.38	-.28*	.22	4.44*	.10*
Verbs	14.37	-.29*	.30*	6.72**	.15**
Certainty	.64	-.35**	-.18	5.75**	.13**
Focus on Past	1.18	-.28*	-.15	3.60*	.74*
Focus on Future	3.68	-.18	.20	2.36	.04
Motion	6.95	.08	.44***	7.92***	.18***
Space	16.27	.03	-.12	.53	-.01
Time	8.81	.07	-.16	.95	-.00
Assent	4.46	-.08	-.23	1.90	.03
Interrogatives	1.29	.07	-.41***	6.53**	.15**
Auxiliary Verbs	6.25	-.24	-.04	1.91	.03
Adverbs	7.73	.03	.06	.14	-.03
Comparatives	1.79	-.15	-.06	.90	-.00
Insight	.37	-.15	-.05	.87	-.00

Note. *$p < .05$, **$p < .01$, ***$p<.001$

Appendix F. *Quadratic – Whole Dialogue*

Measures	*M*	β_{Trial}	$\beta_{Trial\hat{}2}$	β_{Pair}	*F*	*Adj. R2*
Word Count	744.91	-.62***	.46***	-.06	31.27***	.58***
Tentativeness	.79	-.28*	.22+	-.10	3.30*	.10*
Focus on Present	11.05	-.27*	.17	.34**	5.73**	.18**
Prepositions	13.53	-.18+	.12	.54***	10.74***	.31***
Causalities	.46	-.39**	-.00	.20+	4.92**	.15**
Discrepancies	.63	-.30*	.02	.30*	4.62**	.14**
Verbs	14.71	-.31**	.13	.34**	6.15***	.19***
Certainty	.54	-.36**	.06	-.19	4.25**	.13**
Focus on Past	1.32	-.36**	-.09	-.20+	4.46**	.14**
Focus on Future	3.76	-.09	.02	.27*	1.86	.04
Motion	7.33	.10	.10	.54***	9.50***	.28***
Space	15.16	-.03	-.04	.09	.20	-.04
Time	10.51	.27*	3.94	-.13	2.21	.05
Assent	4.80	.02	-.05	-.16	.63	-.02
Interrogatives	1.43	-.01	-.18+	-.53***	9.23***	.28***
Auxiliary Verbs	6.69	-.21+	.07	-.04	1.12	.01
Adverbs	7.71	.04	-.07	-.14	.52	-.02
Comparatives	1.95	-2.66*	.97	-.08	1.77	.03
Insight	.51	-.16	-.07	-.10	.87	-.01

Note. *$p < .05$, **$p < .01$, ***$p<.001$

Appendix G. *Quadratic – Path Maneuvers*

Measures	*M*	β_{Trial}	β_{Trial^2}	β_{Pair}	*F*	*Adj.* R_2
Word Count	235.47	-.57***	.42***	-.00	21.46***	.49***
Tentativeness	.71	-.37**	.23*	-.02	4.82**	.15**
Focus on Present	12.47	-.03	.01	.19	.82	-.01
Prepositions	13.26	-.24*	.04	.42***	6.55***	.20***
Causalities	.69	-.29*	.00	.14	2.40	.06
Discrepancies	.46	-.38**	.06	.21	4.88**	.15**
Verbs	16.05	-.19	.01	.25*	2.35	.06
Certainty	.49	-.12	.21+	-.02	1.33	.01
Focus on Past	1.20	-.30*	-.04	-.21	3.34*	.10*
Focus on Future	3.28	-.10	-.00	.31*	2.44	.06
Motion	7.38	.16	-.02	.29*	2.52	.07
Space	13.29	-.10	.07	.24	1.57	.03
Time	8.51	.32**	.10	.02	2.68	.05
Assent	5.21	.05	-.10	-.18	.96	-.00
Interrogatives	1.54	-.02	-.19	-.38**	4.54**	.14**
Auxiliary Verbs	7.18	-.09	.05	-.09	.41	-.03
Adverbs	1.20	.07	-.09	-.02	.28	-.03
Comparatives	1.66	-.18	.05	.04	.77	-.01
Insight	.51	-.14	-.05	-.06	.54	-.02

Note. *$p < .05$, **$p < .01$, ***$p<.001$

Appendix H. *Quadratic – Manner Maneuvers*

Measures	*M*	β_{Trial}	β_{Trial^2}	β_{Pair}	*F*	*Adj. R*2
Word Count	215.23	-.49***	.33**	.04	11.18***	.32***
Tentativeness	.87	-.05	.14	-.03	.51	-.02
Focus on Present	10.23	-.20	.08	.34**	4.03*	.12*
Prepositions	11.01	-.12	.01	.39**	4.18**	.13**
Causalities	8.33	-.26*	.00	.05	1.51	.02
Discrepancies	.79	-.13	-.04	.31*	2.66	.07
Verbs	14.05	-.22+	.09	.34**	4.34**	.13**
Certainty	.59	.05	.10	-.11	.53	-.02
Focus on Past	1.37	-.28*	-.02	-.04	1.84	.04
Focus on Future	4.17	-.05	.12	.09	.53	-.02
Motion	8.33	-.04	.16	.48***	7.17***	.22***
Space	14.75	-.06	-.07	.18	.86	-.01
Time	13.40	.10	.05	-.10	.48	-.02
Assent	5.38	.12	.07	.14	.04	-.01
Interrogatives	1.46	.02	-.21+	-.40***	5.18**	.16**
Auxiliary Verbs	6.99	-.11	.01	-.05	.32	-.03
Adverbs	7.62	-.09	-.03	-.25*	1.63	.03
Comparatives	2.28	-.23+	.07	-.20	2.29	.06
Insight	.67	-.04	-.07	-.03+	.15	-.04

Note. *$p < .05$, **$p < .01$, ***$p<.001$

Appendix I. *Quadratic – Combination Path and Manner Maneuvers*

Measures	*M*	β_{Trial}	β_{Trial^2}	β_{Pair}	*F*	*Adj. R*2
Word Count	295.09	-.55***	.44***	-.13	21.56***	.49***
Tentativeness	.70	-.28*	.21+	-.05	2.93*	.08*
Focus on Present	10.90	-.23+	.20	.24*	3.55*	.11*
Prepositions	15.70	.11	.10	.40***	4.67**	.14**
Causalities	.38	-.28*	.07	.22	3.03*	.09*
Discrepancies	.56	-.40***	-.00	.22	5.36**	.17**
Verbs	14.37	-.29*	.15	.30*	5.13**	.16**
Certainty	.64	-.35**	.06	-.18	3.87*	.12*
Focus on Past	1.18	-.28*	-.08	-.15	2.53	.07
Focus on Future	3.68	-.18	-.14	.20	2.00	.04
Motion	6.95	.08	.09	.44***	5.44**	.17**
Space	16.27	.03	-.07	-.12	.45	-.03
Time	8.81	.07	-.22+	-.16	1.70	.03
Assent	4.46	-.08	-.01	-.23	1.25	.01
Interrogatives	1.29	.07	-.05	-.41***	4.36**	.13**
Auxiliary Verbs	6.25	-.24+	.04	-.04	1.28	.01
Adverbs	7.73	.03	-.10	.06	.32	-.03
Comparatives	1.79	-.15	.07	-.06	.71	-.01
Insight	.37	-.15	.13	-.05	.93	-.00

Note. *$p < .05$, **$p < .01$, ***$p<.001$

Appendix J. *Logarithmic – Whole Dialogue*

Measures	*M*	β_{Trial}	$\beta_{Log(Trial)}$	β_{Pair}	*F*	*Adj. R2*
Word Count	744.91	1.41***	-2.10***	-.06	41.64***	.65***
Tentativeness	.79	.67	-.98*	-.10	3.67*	.11*
Focus on Present	11.05	.47	-.76	.34**	6.05**	.19**
Prepositions	13.53	.27	-.46	.54***	10.64***	.31***
Causalities	.46	-.38	-.00	.20	4.91	.15**
Discrepancies	.63	-.24	-.06	.30*	4.61**	.14**
Verbs	14.71	.20	-.52	.34**	6.15***	.19
Certainty	.54	-.19	-.18	-.19	4.21**	.13**
Focus on Past	1.32	-.88	.53	-.20	4.78**	.15**
Focus on Future	3.76	-.03	-.07	.27*	1.86	.04
Motion	7.33	.43	-.34	.54***	9.40***	.28***
Space	15.16	-.15	.12	.09	.19	-.03
Time	10.51	.05	.33	-.13	2.14	.05
Assent	4.80	-.25	.27	-.16	.68	-.02
Interrogatives	1.43	-.61	.63	-.53***	8.91***	.27***
Auxiliary Verbs	6.69	.15	-.38	-.04	1.22	.01
Adverbs	7.71	-.02	.06	-.14	.43	-.03
Comparatives	1.95	.24	-.51	-.08	1.97	.04
Insight	.51	-.34	.19	-.02	.83	-.01

Note. *$p < .05$, **$p < .01$, ***$p<.001$

Appendix K. *Logarithmic – Path Maneuvers*

Measures	*M*	β_{Trial}	$\beta_{Log(Trial)}$	β_{Pair}	*F*	*Adj. R2*
Word Count	235.47	1.25***	-1.89***	-.00	25.98***	.54***
Tentativeness	.71	.42	-.81	-.02	4.40**	.14**
Focus on Present	12.47	.09	-.12	.19	.84	-.00
Prepositions	13.26	-.15	-.09	.42***	6.51***	.20
Causalities	.69	-.34	.05	.14	2.41	.06
Discrepancies	.46	-.10	-.29	.21	4.12**	.15**
Verbs	16.05	-.12	-.07	.25*	2.36	.06
Certainty	.49	.42	-.56	-.02	.74	-.01
Focus on Past	1.20	-.64	.35	-.21	3.51*	.10*
Focus on Future	3.28	-.11	.00	.31*	2.45	.06
Motion	7.38	.11	.06	.29*	2.51	.07
Space	13.29	.01	-.11	.24	1.49	.02
Time	8.51	.64	-.32	.02	2.60	.07
Assent	5.21	-.29	.35	-.18	.93	-.00
Interrogatives	1.54	-.70	.70	-.38	4.38**	.13**
Auxiliary Verbs	7.18	.21	-.31	-.09	.48	-.02
Adverbs	1.20	-.01	.08	-.02	.12	-.04
Comparatives	1.66	-.01	-.18	.04	.77	-.01
Insight	.51	-.36	.23	-.06	.55	-.02

Note. *$p < .05$, **$p < .01$, ***$p<.001$

Appendix L. *Logarithmic – Manner Maneuvers*

Measures	*M*	β_{Trial}	$\beta_{Log(Trial)}$	β_{Pair}	*F*	*Adj.* R_2
Word Count	215.23	1.00*	-1.54***	.04	13.29***	.36***
Tentativeness	.87	.68	-.75	-.03	.84	-.01
Focus on Present	10.23	.11	-.32	.34**	4.02*	.12*
Prepositions	11.01	.01	-.14	.39**	4.22**	.13**
Causalities	8.33	-.26	.00	.05	1.51	.02
Discrepancies	.79	-.38	.26	.31*	2.73	.07
Verbs	14.05	.05	-.30	.34	4.25**	.13**
Certainty	.59	.40	-.36	-.11	.49	-.02
Focus on Past	1.37	-.53	.26	-.04	1.94	.04
Focus on Future	4.17	.28	-.34	.09	.36	-.03
Motion	8.33	.58	-.63	.48***	7.15***	.22***
Space	14.75	-.16	.10	.18	.76	-.01
Time	13.40	.24	-.15	-.10	.45	-.03
Assent	5.38	.27	-.15	.14	.04	-.01
Interrogatives	1.46	-.72	.77	-.40***	5.02**	.16**
Auxiliary Verbs	6.99	-.15	.04	-.05	.32	-.03
Adverbs	7.62	-.14	.04	-.25*	1.61	.03
Comparatives	2.28	.12	-.36	-.20	2.39	.06
Insight	.67	-.25	.22	-.03	.11	-.04

Note. *$p < .05$, **$p < .01$, ***$p<.001$

Appendix M. *Logarithmic – Combination Path and Manner Maneuvers*

Measures	*M*	β_{Trial}	$\beta_{Log(Trial)}$	β_{Pair}	*F*	*Adj. R2*
Word Count	295.09	1.37***	-1.99***	-.13	27.44***	.55***
Tentativeness	.70	.59	-.90	-.05	3.11*	.09*
Focus on Present	10.90	.64	-.90	.24*	3.94*	.12*
Prepositions	15.70	.50	.78	.40***	4.70**	.15**
Causalities	.38	.15	-.44	.22	3.25*	.09*
Discrepancies	.56	-.47	.08	.22	5.37**	.17**
Verbs	14.37	.34	-.66	.30*	5.27**	.16**
Certainty	.64	-.13	-.23	-.18	3.87*	.12*
Focus on Past	1.18	-.69	.42	-.15	2.65	.07
Focus on Future	3.68	-.59	.43	.20	1.83	.04
Motion	6.95	.37	-.30	.44***	5.39**	.17**
Space	16.27	-.22	.26	-.12	.44	-.03
Time	8.81	-.76	.85	-.16	1.70	.03
Assent	4.46	-.15	.08	-.23	1.25	.01
Interrogatives	1.29	.06	.01	-.41***	4.28**	.13**
Auxiliary Verbs	6.25	.07	-.32	-.04	.02	1.40
Adverbs	7.73	-.13	.16	.06	.13	-.04
Comparatives	1.79	.27	-.44	-.06	.87	-.01
Insight	.37	.34	-.51	-.05	.94	-.00

Note. *$p < .05$, **$p < .01$, ***$p<.001$

References

Anderson, R. C., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. (1977). Frameworks for comprehending discourse. *American educational research journal*, *14*(4), 367-381.

Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, UK: Cambridge University Press.

Bartlett, F.C. (1928). Social constructiveness: Pt. 1. *British Journal of Psychology*, *18*, 388-391.

Chomsky, N. (1953). Systems of Syntactic Analysis. *The Journal of Symbolic Logic*, *18*(3), 242-256.

Chomsky, N. (1970). Remarks on nominalization. In: R. Jacobs and P. Rosenbaum (eds.) Reading in English Transformational Grammar, 184-221. Waltham: Ginn.

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

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge: Cambridge University Press.

Doumas L.A.A., & Hummel J.E. (2005). Approaches to modeling human mental representations: What works, what doesn't, and why. In J. K. Holyoak & R. G. Morrison (Eds.), The Cambridge Handbook of Thinking and Reasoning (82-100). New York: Cambridge University Press.

Jackendoff R., (1991). Parts and boundaries. *Cognition, 41,* 9-45.

Levin B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago, IL: The University of Chicago Press.

Livingston, E. (1987). *Making sense of ethnomethodology*. New York, NY: Routledge & Kegan Paul.

Mackenzie, C.F., Garofalo, E., Shakelford, S., Shalin, V., Pugh, K., Chen, H., Puche, A., Pasley, Sarani, B., Henry, S. & Bowyer, M. (2015). Using an individual procedure score before and after the advanced surgical skills exposure for trauma course training to benchmark a hemorrhage-control performance metric. Journal of Surgical Education, July, pii: S1931-7204(15)00161-0. doi: 10.1016/j.jsurg.2015.06.009.

Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M.E. (2015). Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

Schank, R. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, *3*, 552-631.

Schank, R.C. & Abelson, R. P. (1988). Conceptual dependency theory. In A. Collins and E. E. Smith (Eds.), Readings in cognitive science: A perspective from psychology and artificial intelligence (195-200). Palo Alto, CA: Morgan Kaufmann Publishers, Inc.

Shah, C.S., Shalin, V.L., Pugh, K., Agandi, L., Puche, A., & Tisherman, S. (in press). Communication patterns in remote, interactive surgical training. *Proceedings of the human factors and ergonomics annual meeting.*

Shalin, V. L., Geddes, N. D., Bertram, D., Szczepkowski, M. A., & DuBois, D. (1997). Expertise in dynamic, physical task domains. *Expertise in context* (pp. 195-217).

Talmy, L. (2000). *Toward a cognitive semantics*. Volume I: Concept structuring systems, i-viii, 1-565. Cambridge, MA: MIT Press.