RESILIENT FEATURES OF RE-EMERGING DYADIC COMMUNICATION SYSTEMS IN AN INTERACTIVE VIRTUAL ENVIRONMENT

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MURAT ULUBAY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF COGNITIVE SCIENCE

JANUARY 2013

RESILIENT FEATURES OF RE-EMERGING DYADIC COMMUNICATION SYSTEMS IN AN INTERACTIVE VIRTUAL ENVIRONMENT

Submitted by **MURAT ULUBAY** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in the Department of Cognitive Science**, **Middle East Technical University by**,

Prof. Dr. Nazife Baykal _____
Director, Informatics Institute

Prof. Dr. Cem Bozşahin _____
Head of Department, Cognitive Science

Assist. Prof. Dr. Annette Hohenberger _____
Supervisor, Cognitive Science, METU

Examining Committee Members

Assist. Prof. Dr. Murat Perit Çakır _____
COGS, METU

Assist. Prof. Dr. Annette Hohenberger _____
COGS, METU

Prof. Dr. Cem Bozşahin _____
COGS, METU

Prof. Dr. Kürşat Çağıltay _____
CEIT, METU

Assist. Prof. Dr. Erol Özçelik _____
CENG, ATILIM UNIVERSITY

Date: 21.01.2013

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Murat Ulubay

Signature :

ABSTRACT

RESILIENT FEATURES OF RE-EMERGING DYADIC COMMUNICATION SYSTEMS IN AN INTERACTIVE VIRTUAL ENVIRONMENT

Ulubay, Murat

Ph.D., Department Of Cognitive Science

Supervisor: Assist. Prof. Dr. Annette Hohenberger

February 2013, 193 pages

This study mainly focuses on the emergence and utilization of communication systems in the context of joint action where collective cognitive activity is required. Dyads are given an instant messaging medium of communication where only a limited number of characters and symbols can be used for information exchange in order to collaborate on common tasks of finding objects, in a network-based interactive virtual environment (ActiveWorlds), a 3D, multi-agent, virtual reality platform. The restrictions on communication and the requirement of collaboration facilitated the creation of a lexical inventory and a minimalistic communication system, a compressed version of dyads' shared Natural Languages (NLs). Across eight experimental sessions, two manipulations are made in order to study their effects on parameters on 4 levels of analysis: (1) Quantitative, (2) Syntactic Complexity, (3) Lexical Category and (4) Speech Act Category. The two interventions are (1) increasing the number of targets from one to two after the first three experiments, and (2) administering a two months break between the 6^{th} and 7^{th}-8^{th} experiments.

Increased number of target objects influenced the quantitative parameters that are related to the amount of communication as well as the use scores of lexical, syntactic, and speech act categories; however, the use ratios of several parameters were resilient

under this manipulation and rather showed different trends of change characterizing the development of the system towards a more mature state in accordance with the demands of the task structure. The opposing trends of increasing use ratio of *Assertive* and decreasing use ratio of *Directive Speech Acts* and decreasing use ratios of the *Type/Token Number of Lexical Items* in a session, the *Number of New Lexical Items* in a session and increasing ratio of *Turn Success* are also characteristics of this maturation.

The break administered between the 6th and 7th experimental session did not cause any decay in the acquired skills of using the emerged communication system. The previously negotiated strategies and acquired skills of communication as well as the trends of the use ratios of parameters were resilient.

The qualitative analysis of the developing communication system revealed several strategies, including compression of NL words into new lexical items, exploiting the redundancy of characters of written words, and iconicity and indexicality of given symbols.

The main drivers of the development of the new communication system appeared to be the processes of integration of communicative with behavioral action. The cognitive capacities enabling this integration and the comprehension of the utterances in the new system is explained by the Cognitive and Communicative Principles of Relevance that are attributed to a comprehension sub-module of a mind-reading module of the human cognitive system.

Keywords: Emergence of Communication, Joint Action, Extended Cognition, Collective Cognition, Collaborative Action, Relevance Theory.

ÖZ

SANAL GERÇEKLİK ORTAMINDA OLUŞAN DİYADİK İLETİŞİM SİSTEMLERİNİN KALICI ÖZELLİKLERİ

Ulubay, Murat
Doktora, Bilişsel Bilimler Bölümü
Tez Yöneticisi: Yrd. Doç. Dr. Annette Hohenberger

Şubat 2013, 193 sayfa

Bu çalışma kolektif bilişsel etkinliğin gerektiği durumlarda iletişim sistemlerinin ortaya çıkışına, gelişimine (emerjans) ve kullanımına odaklanmaktadır. Bilişsel Bilim alanındaki benzer çalışmalarla süreklilik içinde, iletişim sistemlerinin oluşum ve gelişimi, katılımcıların dilsel etkinliğinin kısıtlandığı bir deneysel ortamda incelenmiştir. Çiftler halinde gruplanmış katılımcılara kısıtlı sayıda harfler ve semboller ile kullanılan bir anlık mesajlaşma arayüzü ve ActiveWorlds adlı çok etmenli, 3 boyutlu sanal gerçeklik platformunda oluşturulmuş ortamda nesne arama/bulma gibi iletişim gerektiren ve işbirliğine dayalı ortak görevleri içeren bir deneysel ortam sunulmuştur. Kısıtlanmış iletişim imkânı ve işbirliği gereksinimi sayesinde çiftlerin ortak doğal dillerinin daraltılmış bir türü olan minimalist bir iletişim sistemi ve bu sisteme has sözcük dağarcığının oluşması mümkün olmuştur. İlk altısı aynı hafta içine yapılan ardışık sekiz deney oturumu boyunca dört değişik analiz düzeyindeki parametreler üzerindeki etkilerini incelemek amacıyla iki adet müdahalede bulunulmuştur. İlk üç deney boyunca tek olan hedef sayısı dördüncü deneyde ikiye çıkarılmış (1), ve altıncı deneyden sonra yedinci ve sekizinci deneylere kadar 2 ay ara verilmiştir (2). Dört analiz düzeyi ise, Niceliksel Değişkenler, Sentaktik Karmaşıklık, Sözcük Kategorileri ve Söz-eylem Kategorilerindeki parametreleri içermektedir.

Hedef nesne sayısındaki artışın iletişim miktarı ile ilişkili *niceliksel parametreler* üzerinde etkili olduğu görülmüştür (bir deney oturumundaki *Konuşma Sırası Sayısı* gibi). Bu manipülasyonun diğer analiz düzeyi kategorilerdeki *kullanım miktarı* parametrelerini de etkilediği görülmekle beraber, bu üç analiz düzeyindeki görece *kullanım oranlarını* betimleyen parametrelerin değişim trendleri üzerinde etkisinin asgari düzeyde kaldığı veya görülmediği tespit edilmiştir. Bu ikinci tür *kullanım oranı* parametrelerindeki değişim trendlerinin, iletişim sisteminin görev gereksinimleri doğrultusunda olgunlaşma ve yetkinleşme yönündeki gelişimini karakterize eden değişimler olduğu görülmüştür. *Saptayıcı-Betimleyici* (Constative/assertive) ve *Yönlendirici* (directive) *söz-eylemlerin kullanım oranları*ndaki zıt değişim trendleri, bir oturumda kullanılan Çeşit/Örnekçe sözcük sayısı oranı *değişkenindeki* yükselme eğilimi, Her *bir* Oturumda Türetilen Yeni Sözcük Sayısı değişkenindeki azalma eğilimi, yükselen Başarılı Konuşma Sırası *Oranı gibi* eğilimler bu yetkinleşme sürecini karakterize etmektedir.

İkinci manipülasyon olan, altıncı deneyden sonra verilen iki aylık aranın, sistemin, katılımcıların geliştirdiği iletişimsel bir kabiliyet olarak kalıcılığına ve sisteme özel bilgilerin (söz dağarcığı gibi) kalıcılığına bir etkisi olmadığı gözlenmiştir. Yukarıda bahsi geçen ve gelişimi betimleyen genelde kullanım oranı cinsinden olan parametrelerdeki değişim eğilimleri de bu aradan etkilenmemiştir.

Oluşan iletişim sistemlerinin gelişiminin niteliksel çözümlemesi ise sözcük dağarcığına mevcut doğal dillerden mevcut kısıtlara sıkıştırma yoluyla yeni sözcük ekleme, mevcut karakterlerin, simgelerin ikonik/gösterimsel(indexicality) özelliklerinden faydalanma, tek karakterli kelimelerden oluşan telegrafik diller kurma gibi stratejileri açığa çıkarmıştır. Tüm nicel ve nitel parametreler üzerinden iletişim sistemlerinin gelişimi incelendiğinde, etken süreçlerin iletişimsel ve davranışsal edimlerin bütünleşmesi ve koordinasyonu olduğu testpit edilmiştir. Bu bütünleşmeyi ve oluşan iletişim sistemi içinde anlamları kavramayı mümkün kılan bilişsel kapasiteler ise Bilişsel ve İletişimsel Bağıntı (relevance) çalışma prensiplerine göre işleyen, insan bilişsel sisteminin zihin okuma modülüne tabi, kuramsal bir kavrama alt-modülüne atfedilmiştir.

Anahtar Kelimeler: iletişim sistemlerinin oluşumu, ortaklaşa eylem, bilişsel bilim, söz-eylem, bağıntı kuramı, kolektif bilişsellik.

To My Family

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation and gratitude to my supervisor Annette Hohenberger for her guidance, advices, criticism and continuous support, positive and kind attitude and encouragement throughout the study. If it wasn't for her, I would not be at this point.

I must mention the significance of the constructive, enjoying attitude of the participants of the experiments. They have devoted a substantial time of their daily life for participating in this extended experimental study in a voluntary basis, I am grateful to all.

I would like to express my deepest gratitude to my dear Hilal, who helped me a lot not just in the general form of support but also as in the laborious phases of data analysis and writing the thesis.

I would like to thank my dear friends Fırat Soylu, Esra Can, Zelal Özdemir, Ulaş Başar Gezgin, Reha Özalp and Can Acan who had supported me a lot during my PhD study from the very earlier stages, with their sharing of my excitement, intellectual discussions and various other ways.

Finally, I would like to express my deepest appreciation to my co-workers Ulaş Canatalı and Gül Çalışır Acan for their endless support, and Feride Erdal for her understanding and patience during this process.

TABLE OF CONTENTS

ABSTRACT iv
ÖZ vi
DEDICATION viii
ACKNOWLEDGEMENTS ix
TABLE OF CONTENTS x
LIST OF TABLES xv
LIST OF FIGURES xvi
ABBREVIATIONS xx
CHAPTER
1. INTRODUCTION 1
1.1. Problem Statement (s): 4
1.2. Significance of the study 4
1.3. Organization of the Dissertation: 6
2. LITERATURE REVIEW 7
2.1. Pragmatics, Cognition and Communication 7
2.1.1. Relevance: Communication and cognition 9
2.1.2. Implementation of Relevance Theory: 11
2.2. Computational Models/Multi-Agent Studies for Emergent Communication and Semiotic Dynamics 14
2.3. Cognitive mechanisms of Joint Action and Communication 17
2.4. Emergence of communication from adaptive interaction; definitions of communication, action/communicative action distinction 20

2.4.1. Causal definition communication 20
2.4.2. Sensory motor interaction as communication 24
2.4.3. Action/Communicative Action Distinction in Contemporary Studies on the subject of Emergence of Communication: 25
2.5. Galantucci's Study on emergence of communication: 32
2.6. DISCUSSION ON THE PRESENTED STUDIES ON THE EMERGENCE OF COMMUNICATION 37
3. METHODS AND EXPERIMENTS 39
3.1. Pilot Experiment 1: Google Earth Experiment 40
3.2. Pilot Experiment 2: ActiveWorlds Environment 43
3.3. Comparison with Galantucci's Experiment 48
3.4. The parameters and the modified and finalized design of the present experiments 49
3.5. General Hypothesis and Research Parameters 50
3.6. MAIN EXPERIMENTS and Data Processing for DESCRIPTIVE ANALYSIS: 51
3.6.1. Experimental Design 52
3.6.2. Data recording and Collection: 56
3.7. Data Analysis: 57
3.7.1. Quantitative and Temporal Analysis: 57
3.7.2. Parametric Statistics: Repeated Measures ANOVA 62
3.7.3. Non-parametric Statistics: Friedman's Test 65
3.7.4. Lexical Category Analysis 66
3.7.5. Syntactic Complexity 72
3.7.6. Speech Act Analysis 74
3.8. Groups of Hypotheses on the Four Levels of Analysis 79
3.8.1. The overall change of a parameter across the eight experiments 79
3.8.2. The effect of increasing the target number in the 4th experiment 79

3.8.3. The effect of the two months gap between the 6th and last two experiments (btw. 6-7,8th) 80
3.9. Production of Actual Hypotheses 81
3.10. List of Hyphotheses 82
4. THE DESCRIPTIVE AND STATISTICAL ANALYSIS OF THE EXPERIMENTAL DATA AND QUALITATIVE RESULTS 90
4.1. Quantitative and Temporal Analysis: 93
4.1.1. TIME 93
4.1.2. Total number of turns: 94
4.1.3. Turn Success Ratio 97
4.1.4. Token number of lexical items: 99
4.1.5. Number of New Lexical Items 103
4.1.6. Type Number of Lexical Items: 106
4.1.7. TYPE over TOKEN number of lexical items ratio: 108
4.2. Syntactic Complexity Analysis: 113
4.2.1. Complexity Parameter**:** 113
4.3. LEXICAL CATEGORY ANALYSIS 117
4.3.1. Token Number of Nouns 118
4.3.2. Token Number of Pronouns 121
4.3.3. Token Number of Verbs: 123
4.3.4. Token number of Adjectives and Adverbs 124
4.3.5. Token number of Interjections 126
4.3.6. The Overall Analysis of Lexical Categories 129
4.4. SPEECH ACT CATEGORY ANALYSIS 133
4.4.1. Assertive Speech Acts: 134
4.4.2. Directive speech act category 136
4.4.3. Expressive Speech Acts 139
4.4.4. Commissive Speech Acts 141

4.4.5. The Overall Analysis of Speech Act Categories 143
4.5. General Discussion on the 4 Levels of Parameters 145
4.6. RESULTS OF THE QUALITATIVE ANALYSIS OF THE EXPERIMENTAL DATA . 150
4.6.1. The Compression of NL Utterances into a Given Communication Tool 151
4.6.2. The Parsimonious Single Character Strategy: 152
4.6.3. Substituting the Redundancy 154
4.6.4. The Search in two NL lexicons to find Compressible NL lexical items: Exploiting the Second Language: 156
4.6.5. The Utilization of Written Communication and Instant Messaging Environment: 157
4.6.6. General Cognitive Skills and Traits REQUIRED AND Used to Establish Conventions: 157
4.6.7. Existing Communication Skills of Conversation 160
4.6.8. Is the New System a Compressed NL or are there any Unique Features? 161
5. DISCUSSION AND CONCLUSION 163
5.1. The Communication System 163
5.1.1. The significance of the Results 166
5.2. The Dimensions of Compression 167
5.3. With or without a Dedicated Communication Channel: Does it really matter? 169
5.4. From Action to Communicative action 171
5.5. Participants' Subjective Cognitive Accounts of their Communicative Activity 173
5.6. Limitations of the study: 175
5.7. conclusion 177
REFERENCES 179

APPENDICES 185
APPENDIX A PRESENTATION AND INSTRUCTIONS OF THE EXPERIMENTS 185
APPENDIX B EXPERIMENT TASKS FOR THE 8 EXPERIMENTS 186
APPENDIX C GUIDELINE FOR SPEECH ACT CATEGORY JUDGMENTS 189
CURRICULUM VITAE 192

LIST OF TABLES

Table 1 The 8 experiments' task list 54
Table 2 Sample from post experiment interview records 56
Table 3 Sample of processed communication log 57
Table 4 List of Quantitative Parameters for a Couple's session data 58
Table 5 A sample Quantitative and Temporal Analysis for CP 8 Exp7 59
Table 6 A summary of quantitative parameters for the 8 exps. of CP 8 60
Table 7 The probability of F-statistic for Repeated Measures Anova 64
Table 8 Mauchly's test for sphericity is satisfied $p > 0.05$ 64
Table 9 Saphiro-Wilks test of Normality 66
Table 10 List of lexical items, lexical categorization of words CP 7 69
Table 11 A Sample Table of Token Number of Lexical Items (CP 7) 70
Table 12 A Sample Table of the Normalized Number (Use ratios) of Lexical Categories for (CP 7) 70
Table 13 A Sample Table of Use ratios (Normalized Number) of Nouns 71
Table 14 Syntactic Complexity Score 73
Table 15 Definitions and categories of Speech Acts 75
Table 16 Speech act Categories for Couple 6 across the 8 experiments 77
Table 17 Assertive Speech Acts across the 8 experiments for 8 couples 78
Table 18 Number of lexical types and tokens and Type/Token ratios 108
Table 19 Sample Communication Transcription of Couple 8 152
Table 20 Sample Communication Transcription of Couple 8 154
Table 21 Sample Communication Transcription of Couple 7 155

LIST OF FIGURES

Figure 1 Classification of interaction types (Oliphant, 1997, p. 10)..... 22
Figure 2 Causal Definition of Communication (Oliphant, 1997, p. 12)..... 23
Figure 3 The experimental environment in Scott-Phillips et al. study (2009, p. 228)27
Figure 4 Typical Emergent System (Scott-Phillips et al., 2009, p. 228)..... 29
Figure 5 Gestures as frozen fragment of behaviour 30
Figure 6 The communication medium in Galantucci's experiment..... 33
Figure 7 The maze and the rooms..... 33
Figure 8 Some communication strategies (Galantucci, 2005, p. 740) 34
Figure 9 First-person view of the ActiveWorlds environment..... 44
Figure 10 Second-person view of the participating agent..... 45
Figure 11 General view of the 3D maze 45
Figure 12 ActiveWorlds context of for the sample communication..... 47
Figure 13 A sample graph of a quantitative parameter: New Lexicals.....61
Figure 14 The Mean value of normalized values (use ratios) for nouns..... 72
Figure 15 The Classical categories of speech acts..... 74
Figure 16 The Number of Speech Acts for Couple 6 77
Figure 17 Average Number of Assertive Turns 78
Figure 18 Average Numbers of Turns for 8 Couples 91
Figure 19 Average Experimental Session Time for 8 Couples..... 93
Figure 20 Experimental Session Time for 8 Couples across 8 experiments..... 94
Figure 21 Average Number of Turns of all Couples across 8 experiments 97
Figure 22 Turn Success Ratio for the 8 couples across the 8 experiments..... 98

Figure 23 Average Turn Success Ratio of 8 couples……..99
Figure 24 Estimated Marginal Means of Token number of Lexical Items for all Couples across Exp4-5-6-7-8 101
Figure 25 Average Token Number of lexical items (across all 8 couples)……..102
Figure 26 Average Token number of lexical items (across all couples except CP1) 102
Figure 27 Number of new lexical items for the 8 couples 103
Figure 28 Estimated marginal means plot for the number of new lexical items in the RMA 104
Figure 29 Estimated marginal means plot for the number of new lexical items in the Repeated Measures ANOVA test (for the 6th, 7th and 8th experiments period) 105
Figure 30 Average Number of new lexical items across the 8 exps 106
Figure 31 Average number of the type number of lexical items 107
Figure 32 Average Number of new lexical items 109
Figure 33 Average Type/Token ratio across the 8 experiments 110
Figure 34 The estimated marginal means plot for the *type over token number of lexical items ratio* 110
Figure 35 The estimated marginal means plot for the *type over token number of lexical items ratio* in the Repeated Measures ANOVA test (for the 4-5-6-7-8th experiments) 111
Figure 36 p1/p2 turn taking ratio of the 8 couples across the 8 exps 112
Figure 37 The estimated marginal means plot for the *turn taking ratio* in the Repeated Measures ANOVA test (for the eight experiments) 113
Figure 38-39 Number of Turns……114
Figure 40 Normalized complexity scores (upper panel) and average normalized complexity scores (lower panel) 115
Figure 41 Estimated marginal means plot for the normalized complexity score in the Repeated Measures ANOVA test 117
Figure 42 Token number of nouns (CP1 is Couple 1, CP2, … CP8) 118

Figure 43 Token Nouns Normalized 119
Figure 44 Token Nouns Average / Nouns Normalized Average 120
Figure 45 Average Number of Pronouns 121
Figure 46 Average Normalized Counts of Pronouns (use ratio) 122
Figure 47 Average Number of Verbs across the 8 experiments 123
Figure 48 Average Normalized Counts of Verbs (use ratio) 124
Figure 49 Average Number of Adj./Adv. Across the 8 experiments 125
Figure 50 Average Normalized Adj./Adv. (Use ratio) 125
Figure 51 Average Token Number of Interjections across the 8 exps 127
Figure 52 Average Normalized Number of Interjections 127
Figure 53 Estimated Means of Normalized Number of Interjections 128
Figure 54 The average token number of lexical categories 129
Figure 55 Average token number of lexical categories across the 8 experiments, interjection excluded 130
Figure 56 The average normalized number (use ratio) of lexical categories across the 8 experiments 130
Figure 57 The average normalized number (use ratio) of lexical categories across the 8 experiments, interjections excluded 131
Figure 58 Assertive Average across the 8 experiments 134
Figure 59 Assertive Normalized Average 135
Figure 60 Number of Directive speech acts for the 8 couples across the 8 experiments 136
Figure 61 Directive Speech Acts Normalized for the 8 couples 137
Figure 62 Average Number of Directive Speech Act Category 137
Figure 63 Average of Directive Speech Acts Normalized 138
Figure 64 Average number of Expressive Speech Acts 139
Figure 65 Average number of Normalized Expressive Speech Acts 140
Figure 66 Number of Commissive Speech Acts 141

Figure 67 Commissive Normalized Speech Acts 142
Figure 68 The Average Normalized Number (use ratio) of Commissive Speech Acts (in Exp4-5-6-7-8)..... 143
Figure 69 Speech Act Categories in average total numbers of 8 couples 144
Figure 70 Speech Act Categories in average (use ratio) of 8 couples..... 144
Figure 71 Generalized graph depicting the typical development of communication systems across the 8 experiments 145
Figure 72 Generalized graph depicting the typical development of communication systems (reproduced here for convenience) 165

ABBREVIATIONS

NL Natural language, human language

RMA Repeated Measures ANOVA

CHAPTER 1

1. INTRODUCTION

Cognition in social context assumes individuals interacting and communicating with other cognitive agents. In other words, language and communicative behavior are prototypical examples of collective cognitive activity (Hutchens and Johnson 2009). The focus on cognition in such contexts brings about a new paradigm of studying cognition. While in the past, research in cognitive science primarily had individuals as a single cognitive system in its focus, this focus is now shifting towards the conception of a group of cognitive agents and to the processes while they are interacting in order to accomplish of a common task and functioning in collaborative and coordinated joint action (Galantucci2005; Steels, 2006, Sebanz and Knoblich 2006). Evidently, the key concept for this interaction and coordination is “communication”. Studies within this new paradigm may vary in terms of their emphasis, for some the utilization of communication is the key point, while some others focus on basic perception-actions links as the building blocks to initiate communication (Wagner et al., 2003).

Usually, natural language (henceforth, NL) supports communication in groups of human individuals. In this context, the main focus hitherto had been on issues like the use, emergence and acquisition of natural language. However, now there is a new trend that emphasizes the interpersonal interaction and the cognitive abilities that facilitates the creation of new communication systems among

individuals in a proper context and under specific conditions. Unlike being exposed to an established natural language, now, the acquisition and creation of new communication systems receive attention, like the spontaneous invention of sign language by deaf children in an impoverished linguistic environment (Goldin-Meadow 2004), or the emergence of Nicaraguan sign language in a school community of deaf students (Senghas, Kita, and Özyürek 2004).

This new approach tries to explain the emergence of natural language by not focusing on the individual speaker or the biologically and specifically designed brain or built-in mental faculties of the speaker, but rather by focusing on the "spoken" (or, signed) language itself (Kirby, 2006, 2000c). This new approach considers language in evolutionary terms, not merely in terms of biological neural evolution of humans. Language and emergence of language as a communication system is studied as a distinct entity replicating itself within human populations. The main approach for these is "the complex adaptive systems" view where "language is a novel system of information transmission" (Kirby 2002, 188). The complex adaptive systems approach uses computational or mathematical models of populations of individuals. These models or simulations try to mimic and examine evolutionary paths of symbolic systems not limited to natural language (Steels, 2006).

This approach is fruitful for research which is not concerned merely with the evolution of natural language but also with the emergence of simpler communication systems which support the collaboration of multiple cognitive agents for a specific task or a limited number of tasks. For a study of real human agents where the collective cognitive ability of forming a communication system is under scrutiny as in the present study, the findings of complex adaptive systems research may be both guiding and be tested for verification.

Closer to the aim of the present study is a research trend of studying the emergence of communication in experimental contexts with human participants (Galantucci 2005, Scott-Phillips, Kirby, and Ritchie, 2009). All of these studies clearly

have the limitation of working with agents who already have some communicative competence and language system. These studies try to gain insight into the emergence of novel communication systems by limiting these existing capabilities by constraining the linguistic actions and the kinds of possible actions in the experimental environment. Under these restrictions participants are exposed to an experimental (generally virtual) environment where they are expected to accomplish certain tasks that may require cooperation/collaboration and communicative action. This need for communicative action is expected to trigger the creation of a set of initial simple conventions within the experimental setting and within a provided communication channel that restricts the use of NL, or that restricts similar naturally available communication opportunities. After a repeated or prolonged exposition of the participants to similar tasks and environments, a communication system is expected to develop gradually. Some experimental settings do not provide any dedicated communicative channel at all (Scott-Phillips, Kirby, and Ritchie, 2009) and rather focus on "signaling the signalhood" in their terms: the very emergence of salient communicative action in comparison to *general non-communicative action* in the environment.

Following these novel developments and trends in the disciplines of Linguistics and Cognitive Science, the present study will concentrate on the emergence of communication systems among collaborating real, human cognitive agents in a dyadic setting, within a computer-based experimental environment and a computer-based communication channel. However, although this study provides a communication channel and an experimental setting that restricts the use of NL or any other already available communication system; it enables and expects a NL-like communication system and lexical inventory to emerge. As the common strategy is to compress existing NL competence within the constraints of the given communication channel, the present study may be dubbed as a study of *re-emergence of a communication system*.

The initial aim of the study design using successive experimental sessions is to force individuals to create a communication system and to study this system's function for and relation with, the cognitive task that the participants will perform collaboratively. Thus, the main goal is to characterize the (re-)emergence of communication systems by means of specified language parameters or features.

1.1. PROBLEM STATEMENT (S):

The initial research questions are as follows:

1. Are individuals capable and if so, to what extent are they capable of creating a new/alternative communication system under specific linguistic and environmental conditions?

2. Are there arbitrary variations among possible communication systems that are expected to emerge or are there resilient features and trends that materialize reliably across these communication systems?

3. During After the emergence of a novel but yet preliminary communication system, how do individuals reach a convention and align their communicative actions in terms of symbols or signs?

4. How is the emerging communication system related to the task at hand, the environmental constraints or features, namely to reach a certain goal in the participants' joint action? In other words, to what degree can these communication systems be considered as an adaptive system that is inherently integrated with/determined by the requirement of a general set of adaptive skills that are constrained by the requirements and necessities of the tasks and environment?

1.2. SIGNIFICANCE OF THE STUDY

The previous studies on emergence of communication systems with human agents used different limitations for the communication channel they provided and various designs for the experimental environment, depending on the phenomena of communication they expected to develop. This was done mostly to see whether any

communicative system emergence at all or not, and, if it did, to see the features of the emerging system depending on the restrictions of the experimental design. For example if you provide a set of tools that allows you to send only very simple and temporally limited graphical messages, the questions may be about the persistence of conventions in the system or alignment to similar conventions by the partners. In such a context the level of development of the emerging communication system has to be preliminary with respect to the real communication systems like NL or other computer based ones (also containing NL), which these participants employ in their real lives. However since the experimental restrictions of this study on the available communication channel allows the creation of a communication system which is much closer to a fully developed (real human communication system like) one, the design of the present study allows studying the following aspects of (re-)emergence of a communication system:

1- We can expect to detect different characteristics of developing communication systems in their preliminary, immature, and later mature stages of development.

2- We can use parameters that can both be applied to real life communication systems, namely, NL and our developing communicative system, examine the observational data in terms of quantitative features on various linguistic levels, such as lexicon, syntax, or speech acts etc., to see the characteristics that can be applied to real life systems [alternatively: “to see what characteristics are operative in real-life systems”].

3- We can expect to see various strategies to overcome the restrictions, strategies that allow for the re-emergence of a communication system, and the cognitive/communicative capabilities employed by the subjects that enable this re-emergence and collaboration for the task fulfillment. This observational information about strategies facilitates

identifying the multi-modal communicative and other adaptive skills (textual, graphic, iconic, deictic, context awareness, embodied etc.) that may be at work in that particular experimental context which are otherwise not visible [invisible].

1.3. ORGANIZATION OF THE DISSERTATION:

There are five chapters in this thesis study. In this first chapter, the focus of the study is introduced and problem statements are presented.

The second chapter provides the literature review where similar studies, their theoretical and experimental approaches and relevant findings are presented and compared.

The third chapter is the method part which comprises experimental design, the relevant assumptions, guidelines, methodological limitations and considerations. This section also contains information about the data gathering and analysis procedures.

The forth chapter is the results section where four sets of parameters (quantitative, lexical category, syntactic complexity, and speech acts) are presented descriptively, statistically analyzed and discussed on the respective parameter level. This chapter also contains the specific discussion of the results of the qualitative analysis of the experimental data.

The fifth chapter is reserved for the general discussion where the quantitative and qualitative results of the several levels of parameters are evaluated in an integrated perspective. Resilient and non-resilient features across the successive experimental sessions and universal as well as individual features across the experimental couples are presented. A conclusion section is included at the end of the fifth chapter.

CHAPTER 2

2. LITERATURE REVIEW

In this chapter, the theoretical background and research findings relevant for emergence of communication studies will be presented. Different aspects of this background are presented as follows:

1- From conventional pragmatics to cognitive notions of processing communicative input: “relevance”
2- Multi-agent simulations and their results
3- Joint action and communication
4- Definitions of communicative action, adaptive development and action/communicative action distinction
5- Galantucci’s experiment on emergence of communication systems.

2.1. PRAGMATICS, COGNITION AND COMMUNICATION

Language and language use is studied in cognitive science mostly maintaining the continuity between the mainstream conceptions of the disciplines of linguistics and psychology: Language as an abstract entity of its own in formal linguistics, language as a separate cognitive module or language and its use as an individual psychological activity (Akmajian, 2001). However, consistent with its interdisciplinary nature, cognitive science developed a new understanding of language (use) as a communication tool between cognitive agents. This approach became a part of the agenda after certain criticisms (Thagard, 2005, p155, p219),

where the computational representational model appeared to be in need of expanding in both social and biological dimensions.

One example of this trend is the idea that communication science can contribute to research on the mind (Stenning et al., 2006, p 14). Following the revision and elaboration of models of communication in pragmatics, i.e. from the *message* model or the *code* model (with the *conduit* metaphor) to the *inference* model (Grice, 1975), (Akmajian, et al., 2001, p 371), we can say that the understanding of the communication process between two agents became a task of understanding the role of the mental capability of *situation and context awareness* and the ability of drawing inferences about the intentions/beliefs of the other agent (Stenning et al., 2006, p 335). This elaboration of communication research and models proceeded in a cognitive direction with the "Cognitive Principle of Relevance" of Dan Sperber. Before giving a synopsis of Sperber's theory, the route from the code model to the inferential model of Grice and then to relevance theory will be outlined

The prototypical model of communication is named [referred to] as code (message) model, where the general assumption is that the [communication] process can be described with the following elements and relations: a speaker, a hearer, and a communication channel. (1) The speaker's intended meaning is encoded into a message by means of language; (2) the communicative process of "uttering" is the conduit (channel) that carries the encoded message; (3) the hearer receives the message, decodes it by means of her linguistic competence, and accesses the representation of the intended meaning as the logical form of the sentence (Wilson and Sperber, 2006, p 606; Akmajian et al., 2002, p 271). This view of encoding-decoding processes is considered to be only partially capturing the communication, because of the fact that "the linguistic meaning recovered by decoding vastly underdetermines the speaker's meaning" (Wilson and Sperber, 2002, p. 3). That is, there are many possible meanings that can be encoded into the

same or similar utterances. So, to facilitate the access to the intended meaning, there is a need of expression and recognition of intentions both by the speaker and the receiver side. This is formulated by Grice as follows when defining sentence meaning and speaker's meaning ($meant_{NN}$), ("NN" in the subscript refers to non-Natural) respectively:

'A meant something by *x*': 'A uttered *x* with the intention of inducing a belief by means of the recognition of this intention.' (Grice 1957 in 1989, p. 220)

2.1.1. RELEVANCE: COMMUNICATION AND COGNITION

Dan Sperber and Deirdre Wilson's work "Relevance: Communication and Cognition" (1995) scrutinizes the notion of relevance in order to focus on and develop the Maxim of Relevance, one of the conversational maxims of Grice's pragmatics, which they claim to be "left relatively undeveloped" (Wilson, 1999). The Gricean Maxim of Relevance is one of his four maxims of inferential communication: Quality (truthfulness), Quantity (informativeness), Relation (relevance) and Manner (clarity). Relevance here is the assumption about the informational content of the message where this new message should add or give information related to the context and subject matter of communication (Levinson, 1983, p. 107). The following perspective of handling relevance as a general cognitive notion is a good example of cognitive examination of communication and focusing on communication in terms of cognitive terms:[alternatively: a good example of how cognitive science approaches communication in its own scientific terms:]

> "In relevance-theoretic terms, any external stimulus or internal representation which provides an input to cognitive processes may be relevant to an individual at some time. According to relevance theory, utterances raise expectations of relevance not because speakers are expected to obey a Co-operative Principle and maxims or some other specifically communicative convention, but because the search for relevance is a basic feature of human cognition, which communicators may exploit." (Sperber and Wilson, 1995, p. 119)

The cognitive principle of relevance claims that the human cognitive system works economically by managing "the relevance of inputs" and "processing efforts to infer relevance" (ibid.). In a context of available assumptions [about the meaning of the input], more relevant input which requires less processing effort has a 'positive cognitive effect', while irrelevant input (with 'negative cognitive effect') does not change the mental state. It is related to a cognitive theory of communication in a way that this universal tendency of maximizing relevance enables individuals "to predict and manipulate the mental states of others." (Sperber, Wilson, 1995, ibid.). This cognitive principle of relevance is formulated into [as] a process or mechanism of pragmatic interpretation "which is ultimately an exercise in mind-reading, involving the inferential attribution of intentions." (Sperber and Wilson, 2002, p. 4) Selecting the most relevant interpretation is only one side of the economy; the other side is to minimize the required effort to make the inference. So the hearer is justified to select the less effort-requiring, or "easy", interpretation, when all other conditions are equal (the degree of relevance or amount of positive cognitive effect).

> "In relevance-theoretic terms, other things being equal, the greater the PROCESSING EFFORT required, the less relevant the input will be. Thus, RELEVANCE may be assessed in terms of cognitive effects and processing effort" (Wilson and Sperber, 2006, p 610)

The procedure of interpreting the speaker's meanings is based on the cognitive principle of relevance as follows:

1- Follow a path of least effort in constructing and testing interpretive hypotheses (regarding disambiguation, reference resolutions, implicatures, etc.).
2- Stop when your expectations of relevance are satisfied (Van der Henst & Sperber, 2006)

Two other important notions of relevance theory are the “communicative principle of relevance” and “optimal relevance”. The cognitive principle of relevance was about the dynamics and processing of relevance, the communicative principle is a built-in assumption that every ostensive communicative stimulus (including linguistic and non-linguistic ones) **refers to a presumption of its optimal relevance**. That is to say, if a stimulus is made apparent enough by the speaker, it’s relevance to the hearer is sanctioned by the speaker by the mere existence of the signal with its being made apparent. This presumption provides a corresponding assumption on the hearer’s side and the facilitation of use of cognitive resources of the hearers for the processing of relevance. The second notion of optimal relevance is about the comparative character of relevance rather than being quantitative (Wilson and Sperber, 2006, p. 612). For the selection of the most appropriate interpretation among several ones, a comparison of optimality is required:

> Optimal relevance:
>
> An ostensive stimulus is optimally relevant to an audience if:
>
> 1- It is relevant enough to be worth the audience’s processing effort;
> 2- It is the most relevant one compatible with the communicator’s abilities and preferences. (Wilson and Sperber, 2006, p. 614)

2.1.2. IMPLEMENTATION OF RELEVANCE THEORY:

These were the basic principles and notions of relevance theory of communication. The second main issue is the implementation of these principles and notions for (i) a better explanatory frame for the observed communicative or pragmatic performance of individuals and (ii) the implementation of these as principles and features of the cognitive system in a modular account of inference-making mechanisms and “in particular of mind-reading”. (Wilson and Sperber, 2006, p. 634)

i-) Relevance and Comprehension:

In several or his works Sperber argues that the cognitive economy of relevance provides a better understanding of comprehension in a communicative context. For example he claims that the Gricean principles of truthfulness and the co-operative principle cannot necessarily deal with the production and comprehension in every utterance context, like irony, metaphor, hyperbole etc.:

On this approach, loose talk, metaphor and hyperbole involve no violation of any maxim, but are merely alternative routes to achieving optimal relevance. Whether an utterance is literally, loosely or metaphorically understood will depend on the mutual adjustment of context, context and cognitive effects in the effort to satisfy the hearer's overall expectation of relevance. (Wilson and Sperber, 2006, p. 24)

ii-) Relevance Theory and Mental Architecture:

These relevance-theoretic explanatory principles that characterize the comprehension of utterances assume a well-known general mental capability, namely theory of mind. Sperber and several co-authors of his studies argue that the application of the general mind-reading abilities – which means lengthy chains of reasoning comprised of reflective conscious activity, of rational constructions of how conversational implicatures might be derived etc. – are not plausible for inferring the speaker's meanings (Sperber and Wilson, 2002, p. 13).

Most theories of mind reading assume that mind reading is not performed by a general reasoning system, but by a dedicated module which includes sub-modules like an Eye Direction Detector, an Intentionality detector, the detection of speakers' meanings from utterances. This dedicated sub-module instantiates the previously discussed notions of the cognitive principle of relevance, the communicative presumption of relevance and optimal relevance as built-in automatic and mostly unconsciously activated features (Sperber and Wilson, 2006, p. 635):

> It is hard to believe that two-year-old children, who fail on regular first-order false belief tasks, can recognize and understand the peculiar multi-leveled representations involved in verbal comprehension, using nothing more than a general ability to attribute intentions to agents in order to explain their behaviour. For these reasons, it is worth exploring the possibility that, *within the overall mind-reading module, there has evolved a specialised sub-module dedicated to comprehension, with its own proprietary concepts and mechanisms* (Italics are mine, M.U.) (cf. also Sperber, 1996, 2000, 2002; Origgi and Sperber, 2000,; Wilson, 2000, Sperber and Wilson, 2002).

Sperber also attempts to verify this assumption of a dedicated sub-module of comprehension by means of conducting some experimental studies and surveying relevant studies. The experimental studies comprise relational reasoning, Wason selection and speech productions tasks (Van der Henst & Sperber 2006).

The cognitive and communicative principles of relevance are candidates for an explanatory framework for the emergence of communication systems and their use. The number and categories of building blocks of a communication system can be explained to be limited by this economy (maximizing relevance) of cognitive systems of agents. These cognitive systems may be described as being based on the cognitive principle of relevance and thus avoid production of unnecessarily many [overabundant] communication tools that increase processing costs. These notions provide explanations for how individuals escape the failure of comprehension of the received message by narrowing down the possible alternatives of inference and by acknowledging the presumed relevance. (Compare Grice – Sperber → Extended embodied distributed cognition view of communication (may be at the final discussion)

2.2. COMPUTATIONAL MODELS/MULTI-AGENT STUDIES FOR EMERGENT COMMUNICATION AND SEMIOTIC DYNAMICS

Some researchers used simulations of multi-agent systems to study emergent communication and semiotic dynamics (Panzasara, 2006; Trafton et al., 2006, p. 275). This approach can be related to the research on communication phenomena with an evolutionary and adaptive point of view and be compared to the relevance theoretic study of pragmatics (which has been examined in the previous section) of existing cognitive capabilities of communication required for comprehension and production. Even that account – assuming a dedicated mind-reading module or a dedicated sub-module for comprehension with specific concepts and mechanisms of comprehension – cannot explain completely how new elements enter the system and become preserved. The multi-agent/simulation approach is considered to be an inescapable since there is no access to the historical events that led to the creation of new elements, development of language and general communicative skills (Hutchins and Johnson, 2009, p. 524):

> All of the simulation models of the emergence of language discussed below assumes that change in the system occurs incrementally in a series of encounters among pairs of agents in a community. How we imagine those encounters sets the stage for nearly everything that follows. All of the models assume that language emerged in the context of inter-agent communication rather than simply in the workings of internal mental processes. It is because of this assumption that models of the emergence of language are considered relevant to collective cognition. (Hutchins and Johnson, 2009, p. 525)

This computational approach includes artificial agents equipped with some basic rules of adoption of new lexical items and forming an inventory of arbitrary words related to some abstract contexts. For example the model called “naming game” was played by a population of agents where an object is picked from the environment/ context by an artificial agent. It then searches its memory to match

and select a word for the object and the second agent adopts this selection of the first agent, if there is no conflict with this in its memory. After several runs of this cycle of innovation, alignment and adjustment, a more similar, shared vocabulary is expected to emerge (Steels, 2006b). The concepts and notions of neural-network-like associative memory or communication systems for virtual agents were the main context of this research but the same results were obtained in human invention and alignment behavior (Steels, 2006a).

Such multi-agent studies are used to compare the results of the model with human semiotic dynamics (Baronchelli, Caglioti, and Steels, 2006). In the area of semiotic dynamics, the dynamics of communicative signs among the population of agents (Kaplan and Steels, 1999) is studied. This is done by observing how populations of humans or agents can establish and share semiotic systems, typically driven by their use in communication (Steels, 2006b). Real-world strategies of real agents are labeling (tagging) and linking different symbolic objects (like photos and articles, Flickr, Wikipedia) and this invention and negotiation of shared semiotic systems emerges by collective cognitive activity of individuals with simple patterns of principles of action. Thus, the multi-agent system analysis is also an area to study human semiotic dynamics within a simulated environment. In semiotic dynamics, human agents' flexibility of adopting other human agents' usage on all levels of language is well studied (Kaplan, 2005) and may explain the emergence of communication systems with real agents.

The simulation approach is considered to be prolific in terms of findings indicating that unanticipated structures can emerge from simple action couplings constrained by basic rules (a particular topology of interaction imposed by the design of the simulation). One example is the alignment of behavior with the other members of the population, such that the population can converge on shared patterns (Hutchins and Johnson, 2009, p. 526). If there is a selection criterion for certain types of action couplings in this process of alignment this specific change in

the system is preserved and propagated across the population. This process of evolution of the system is described by the concept of *modulated positive feedback*: if certain types of action are modulated/filtered (selection mechanism of evolution), the local interaction of individual agents at the microscopic level leads to [the emergence of] a macroscopic structure (Steels, 2007). The use of the above notions and processes as an explanatory frame and the experiments with artificial agents are within the general paradigm of "language as a complex adaptive system".

There is a set of issues that can be considered as criticism for this trend of studies of emergence of communication with artificial agents' experiments. They use predetermined symbolic forms and symbolic meanings in most of the cases; the agent is simply a computer program that uses the conduit metaphor or the code-decode model of communication which does not provide a sense of how the signals or meanings are perceived; and lastly, there is no insight in the origins of symbolic forms (Hutchins and Johnson, 2009, p. 528).

The above criticism is mainly about the fact that the virtual agents are totally disembodied. Experiments with robotic agents introduce a real agent, perceiving and interacting within the environment. In Luc Steels' series of robotic experiments the interaction of other embodied agents, the interaction of motor and visual stimuli provided some of the missing elements (Nolfi, 2005, p. 234). The agent needed and used the information about what the visual scene of the other agent (direction of gaze) is. The direction of motion and similar behavioral cues about other agents are also included in a naming game. The innovations here are: (a) "the separation of dedicated modalities for experiencing the world and communicating about it"; and (b) "the use of real perceptual processes for engaging the world constrains the possible conceptual structures that can emerge" (Hutchins and Johnson, 2009, p. 529). These results show that "artificial systems can handle the symbol grounding problem" (ibid.) when a social view is taken on symbol emergence:

> Pragmatic feedback is crucial for bootstrapping grounded communication. It requires attention sharing, face identification and tracking, gestural recognition, shared task awareness, script execution and recognition, emotion recognition and synthesis, etc. Many of these capabilities have already been demonstrated on robots. (Steels, 2003, p. 311)

2.3. COGNITIVE MECHANISMS OF JOINT ACTION AND COMMUNICATION

A social perspective on cognition requires discarding the assumptions of studying cognition by investigating single individuals. In the case of joint action, investigating cognitive activities of more than one agents and how these are organized into a coordinated action is inevitable. All these approaches take a social perspective on cognition. Some use distributed cognition (Hutchins, 1995), considering the group of individuals as a cognitive system and the primary unit of analysis. “Others consider the cognitive and neural *processes within the boundaries of individual minds*, acknowledging the major force of **interaction** to shape processes of joint action”. (Galantucci and Sebanz, 2009, p. 256)

This investigation focuses on the online interactions and the supporting mechanisms required for joint action. Online interactions require mechanisms for task sharing, joint attention, action observation, action coordination and agency (Sebanz et al., 2006). These mechanisms are not necessarily related to verbal communication processes or may not contain any communicative processes at all. For example some joint action tasks like moving objects together (two persons case) require the alignment of action which contains the online adjustment of behaviour by actively and continuously observing the participants’ action to create coordination. However, a perception-action coupling is possible only by joint goal representation (Sebanz et al., 2006). The need for shared representation is a common requirement for comprehension tasks in a communicative context.

There are also mechanisms which are more apparently related to joint action that comprises communicative action or that are just related to

communicative action. These seem to require dedicated cognitive capabilities. Some are similar to the ones proposed by the cognitive, relevance-theoretic pragmatic approach that has been discussed in the previous sections. These mechanisms contain the general capacity of mind-reading/theory of mind which is the ability of attributing intentions and mental states (knowing, wishing, etc.) to others. Using inferences about the interlocutor's intentions and mental states to adapt one's behavior is the basis of communication-specific mechanisms (*Shintel* and Keysar,2009).

However, such an inferential process can be demanding in terms of both time and cognitive resources and is therefore not plausible (Sperber and Wilson, 1996; *Shintel* and Keysar, 2009). The main argument about the cost of higher level cognitive tasks was that assuming a dedicated sub-module which utilizes the adaptive pragmatic principles (cognitive principle of relevance) works unconsciously and spares the individual from actively reasoning about the mental states and representations of the world of others, inferring their goals, etc. On the other hand, the joint action studies converge on another and partially opposing solution that there are several low-level mechanisms that enable comprehension in communication and coordination in general joint action tasks. For the case of communication, Shintel and Keysar state:

> Interaction provides interlocutors with many cues that can support coordination of meaning, even when they are neither produced intentionally for that purpose nor interpreted as signaling speakers' intention. In many circumstances, interlocutors can take advantage of these cues to adapt their behavior in ways that promote coordination, bypassing the need to resort to deliberative inferential processes (*Shintel* and Keysar, 2009, p. 260).

The main low-level mechanism that is emphasized here is ***interactive alignment***, where two speakers "simultaneously align their representations at

different linguistic levels" (ibid.) and by imitating each other's choices of speech sounds, grammatical preferences, lexical preferences and meanings:

> Crucially, interactive alignment occurs at all linguistic levels, from the phonological and syntactic up to the semantic and the contextual. Additionally, alignment at one linguistic level leads to greater alignment at other levels [44, 47]. For example, 'low-level' alignment of words or grammatical forms can lead to alignment at the critical level of the situation model (i.e. the level at which the speaker and the listener understand that they are referring. (Hasson et al., 2012, p. 117)

A more unified approach is that human agents need to co-adapt communicative and non-communicative behaviors, and integrate their coordinative skills developed for both. Similar to the relevance-theoretic argument of "under-determination of meaning by the utterance"; and in a more positive sense, the advantage of "relying on the implicit information that does not need to be communicated" requires the exploitation and coordination of non-communicative abilities (Nolfi, 2005, p. 232). When you rely on shared implicit information, the load on the communicative content (the utterance) is less (the expressive content of the surface structure). This renders the correct interpretation also easier as long as the implicit information is shared by the hearer. So this reliance facilitates both the production and comprehension aspects of communication. However, sharing or accessing the implicit information requires employing several more complex cognitive abilities. These can be exemplified as: "the ability to predict the sensorimotor consequences of agents' own actions", "the ability to predict changes in the physical and social environment", "the ability to learn from others or to imitate other agents" (Nolfi, 2005, p. 241), and, lastly, "the ability to have access to their own communication acts" (i.e. talking to themselves (Steels, 2003b)).

Another advanced cognitive requirement is being able to use several forms or protocols of communication, as well as communicative content. These protocols are then negotiated and selected during the online interaction ("on the fly"). This is

the ability to modify communication behaviors on the basis of the potential targets of communication acts (Nolfi, 2005, p. 242). For example, a human agent can limit communication acts, filter useful ones, can regulate communication flow (turn taking), and the agent has the ritualized forms of communicative interaction, like *read-back* rituals to guarantee or to receive confirmation from the speaker about the correctness of the reception of the signal.

2.4. EMERGENCE OF COMMUNICATION FROM ADAPTIVE INTERACTION; DEFINITIONS OF COMMUNICATION, ACTION/COMMUNICATIVE ACTION DISTINCTION

For a research initiative concerning the evolution of language or emergence of communication, a discussion about the action/communicative action distinction is critical. Possible definitions may be used for examining the adaptive pressure turning some kinds of actions into communicative ones and how they further develop. The studies vary in terms of handling the problem: there exist philosophical formalisms to define and categorize different types of communicative actions and communication systems (Oliphant, 1997), robotic studies to observe action patterns turning into communicative action (Quinn, 2001), observations of animal communication (hominids) where iconic forms of desired actions are communicated (Hutchins and Johnson, 2009), and experimental studies involving real human subjects who lack a communication channel in the task environment (Scott-Phillips et al., 2009; Galantucci, 2005).

2.4.1. CAUSAL DEFINITION COMMUNICATION

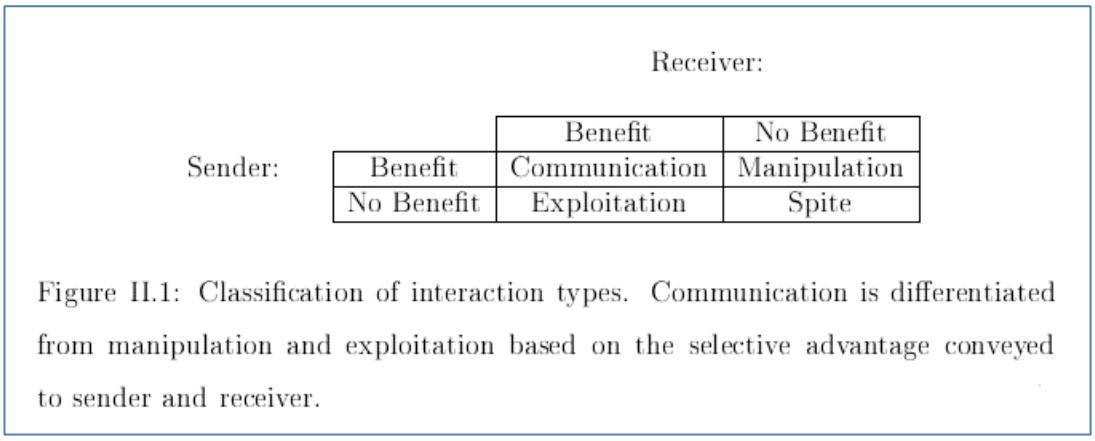
Examples of communication behavior is given by Oliphant (1997)in his study that tries to devise categorical formalisms for communicative actions and then mathematical formalisms to carry out computational simulations of evolution of communication. In this section his categorical formalisms regarding the definition of communication will be presented. The examples of communication behavior

start from a very unusual one, from flowers to pollinators. Insect pheromones, bee dances, alarm calls of primates are more commonsensical ones. The causal or symbiotic relation between flower and pollinator insect is considered as communicative, since the coloration and other visual cues in the flower's appearance directs the insect to the most important zone for pollination. Even though there is no sentience in the plant, the evolutionary/adaptive pressure is considered to be responsible for this kind of structural features to evolve. Finally, "human language" (NL) is listed as the last and most complex form of communication behavior with its combinatorial, syntactic structure. But these features are disregarded and NL's are considered a distinctive subset of the communication systems exemplified previously. The generalizations that apply to these *simple communication systems* are claimed to be applicable to NL as long as these are grounded generalizations on the intersecting features of NL and simple communication systems. Simple communication systems "are those which make use of signals that are independent and discrete. In other words, the signals have no meaning-bearing internal structure, and cannot be combined together to create more complex utterances" (Oliphant, 1997, p. 7).

The controversial example of the flower signaling its own location and the most critical location for the pollination is a good example, because any other examples of animal communication can be considered to be lacking intentionality – except the human communication which requires second-order intentionality (Dennett, 1971), i.e., intention about intention and believes.

Insects may be signaling with pheromones without any intention of communicating, Vervet monkeys' active conscious intention of warning the fellow herd members can be doubted. The Gricean or Dennettian conception of meaning (Cichy, 2005, p. 70) and communication is not adopted by Oliphant; rather, a more causal conception like Lewis and Gower's (1980) is adopted since Oliphant plans to formalize both innate and learned behaviors of communication,

"by being appropriately vague about the locus of a selective advantage, defining communication as **the transmission of a signal or signals between two or more organisms where selection has favored both the production and the reception of the signal(s)**" (1997, p. 10)



		Receiver:	
		Benefit	No Benefit
Sender:	Benefit	Communication	Manipulation
	No Benefit	Exploitation	Spite

Figure II.1: Classification of interaction types. Communication is differentiated from manipulation and exploitation based on the selective advantage conveyed to sender and receiver.

Figure 1 Classification of interaction types (Oliphant, 1997, p. 10)

This preference of defining communication, i.e., employing a notion of the use of the signal without the intention of using it that way seems to be relevant for this study and other emergence of communication studies cited as well. In the following section, we will discuss patterned actions/interactions of artificial agents without any communication channel where they will be exhibiting actions with communicative behavior. There are gradual processes of communicative action becoming distinguished from action in the hominid communication that will be exemplified. And even for mammals with developed brains and human subjects who are very perceptive to the communicative and non-communicative distinction, there are actions which are qualified to be communicative actions after some iterations and realization of an adaptive advantage which is then used with communicative intentions. To be able to grasp these phenomena in their early stages of development this definition may be operational and inspiring. The actual, formal, causal definition of communication and formalization is as follows:

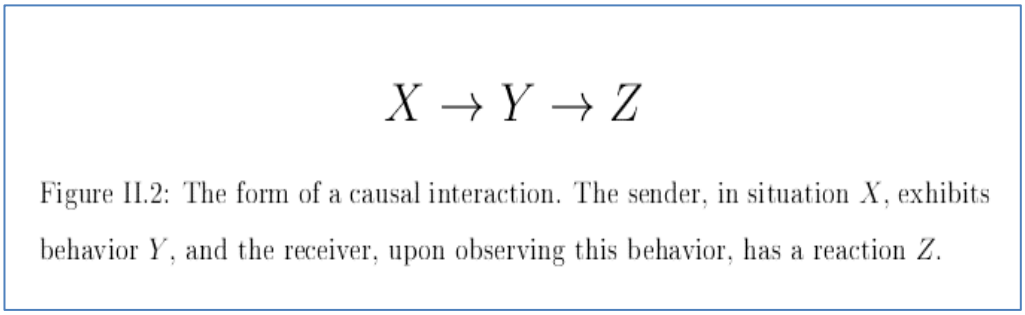


Figure 2 Causal Definition of Communication (Oliphant, 1997, p. 12)

This formalization provides an explanation of the other types of interaction listed in the previous table (Figure 1), like manipulation and exploitation. The causal definition in Figure 2 visualizes the causal chain of the actions of the agent X, X gets (possibly pleasing) behaviour Z of the receiver by the mediation of his action Y. The circularity of the causal definition is apparent: if *selection favors both the production and the reception of the signal* then it is a communicative action. When monkeys see a lion and scream, this behavioral trait has a selective advantage for the population, which is avoiding the predator. Is it the predator avoidance response to the screaming that selects the screaming or does the screaming cause the predator avoidance response and is thus favored by the evolutionary selection processes? The circularity indicates that the traits or screaming and responding to the scream are selected and tuned (exaggerated) over time by a modulated feedback loop (Oliphant 1997, p. 12). The time scale here is the period over generations, i.e., the time of biological evolution; the feedback mechanisms in section 2.2, however, work within generations and the behavioral trait is propagated not by reproduction but interaction among individuals during the lifetime of the generation. To sum up, the causal definition of communication is especially operational when the communicative action is not working in a dedicated channel and the circular causality accounts for the feedback mechanisms for the evolution/emergence of

communicative behaviour. The quote below contains a very good example of these two phenomena:

> The second example involves an innate behavior the threat displays observed in the hostile interactions of many species. One animal, being in an aggressive state, makes a particular display with its body. The other animal, upon observing the display, may withdraw from the interaction. The withdrawal of the second animal is a result of an innate response to the display behavior. This response has been tuned by natural selection because of the association of the display with another animal's state of aggression. The display behavior itself has also been tuned by evolution, being selected for because it has the response that it does namely causing another animal to back down (Oliphant, 1997, p.17).

Again, the threat display may itself be an action of aggression to physically harm the receiver in the beginning but then the display itself can be tuned to a communicative action with the outcome of chasing off without any physical aggression or risk taking required.

2.4.2. SENSORY MOTOR INTERACTION AS COMMUNICATION

In the studies based on simulations with robotic agents, even though there is no dedicated communication channel, certain action patterns turn into and subsequently function as communicative signals. In the experimental environment of Quinn (2001), a set of robotic agents controlled by neural networks are trained under certain task requirements. Each couple of robots evolved as a team where each robot has a motor, wheels allowing rotational and linear motion, an infra-red proximity sensor and a neural network controller. The task is maintaining a distance of 5 cm between the members of the dyad (the sensor range, otherwise they would loose sensory contact) and moving together for a distance of 25 cm in 10 seconds without colliding (Quinn 2001, p. 359). The solution is non-trivial and the only interaction is by means of sensory motor interactions (Nolfi, 2005). This

experimental setting is partially analogous to situations where totally deaf, dumb and blind individuals need to find and lead each other for a certain distance, their only interaction being the tactile feeling, where they will need to avoid harmful collisions, so some of the bodily sensory motor interactions will be performing communicative functions. In the robot simulation the trained neural networks controlling the robots evolved to provide certain action patterns acquiring communicative function after 370 generations. For the encounters where one of the robots is making a reverse rotation when the partner is standing still, or in linear motion the controller makes a change in the direction and the other robot follows to maintain the sensor range. A leader and follower role distribution is provided in this particular form of encounter by means of a particular action coupling. This motor action serves as a signal like an “after you” message (Kirby 2002). Nolfi (2005, p. 234) discusses the character of these actions as signals and communication as follows:

I shall attribute a communication value to all actions or sequences of actions that, by influencing the sensorimotor flow of other agents, enhance the adaptive ability of the group as a whole. The reason why I do not simply call these actions communication acts is that, in addition to a communication value, they might have other functions (e.g. they might allow agents to avoid obstacles, an ability that does not necessarily influence the behaviour of other agents).

2.4.3. ACTION/COMMUNICATIVE ACTION DISTINCTION IN CONTEMPORARY STUDIES ON THE SUBJECT OF EMERGENCE OF COMMUNICATION:

The initial condition of existence of a ready-made channel and tokens (forms of signals) for communication is the “default” for the experimental setting of some of the studies and the use or exploitation of existing communicational skills and natural language competence in that context is not ruled out.

Most of the contemporary research, on the other hand, focuses on the issue of existence or absence of a communication channel. In the case of absence of a pre-determined communication channel, and in the absence of possible forms of signals, the question of "how do potential receivers even know that there is a signal?" is posed in most of these studies (Scott-Phillips et al., 2009). These usually scrutinize the cognitive and interactional processes that create these channels in the first place. Regarding the simulations on emergence of communication systems in the context of embodied agents (like the ones of Steels), Hutchins notes that "*there are two important innovations here: (a) the separation of dedicated modalities for experiencing the world and communicating about it; and (b) the use of real perceptual processes for engaging the world constrains the possible conceptual structures that can emerge*" (Hutchins and Johnson, 2009, p. 529). So the possible dedicated modality for communicating about the world is claimed to be "*inhabiting the same domain of sensorimotor experience as other objects and events in the world*" (Hutchins and Johnson, 2009, p. 532).

As a significant point of discussion and focusing device, the above issues can be considered for a specification of the subject matter: from the study of *an emerging communication system* as a whole and in general, to the narrower scope of observing and studying the *emergence of communicative action*, where individuals acting together as the primary unit of analysis (Galantucci and Sebanz, 2009, p. 256).

The experimental environment and task scenarios in Scott-Phillips' experiment:

Several experimental studies which focused on the emergence of communication are examined for improving the experimental efficiency. In some of these studies the communication channel is pre-determined as in Galantucci (2005), whose study will be elaborated in the next section or as in multi-agent modeling of Steels (2006)). In some others, however, it is not. In the second type of studies like

the one of Scott-Phillips et al. (2009), the participants play an **Embodied Communication Game** (ECG) where the actions required to complete a task (to move in a square divided into four smaller quadrants) at the same time needed to be signaled at some point. The experimental environment is abstract, simple and minimalistic. Simple experimental environments successfully constrain the use of actions in a way that obvious strategies cannot be employed. The second important feature in similar experiments is that the experimental design provides a necessity to communicate as a result of the design and the task. The third important feature is that the production of communicative action is not dependent on users' motivation; rather, the users have to communicate incessantly and each turn in the experiments is completed in a few minutes at most. This provides sessions of experiments with numerous turns played, during which communicative interactions turn into communication systems.

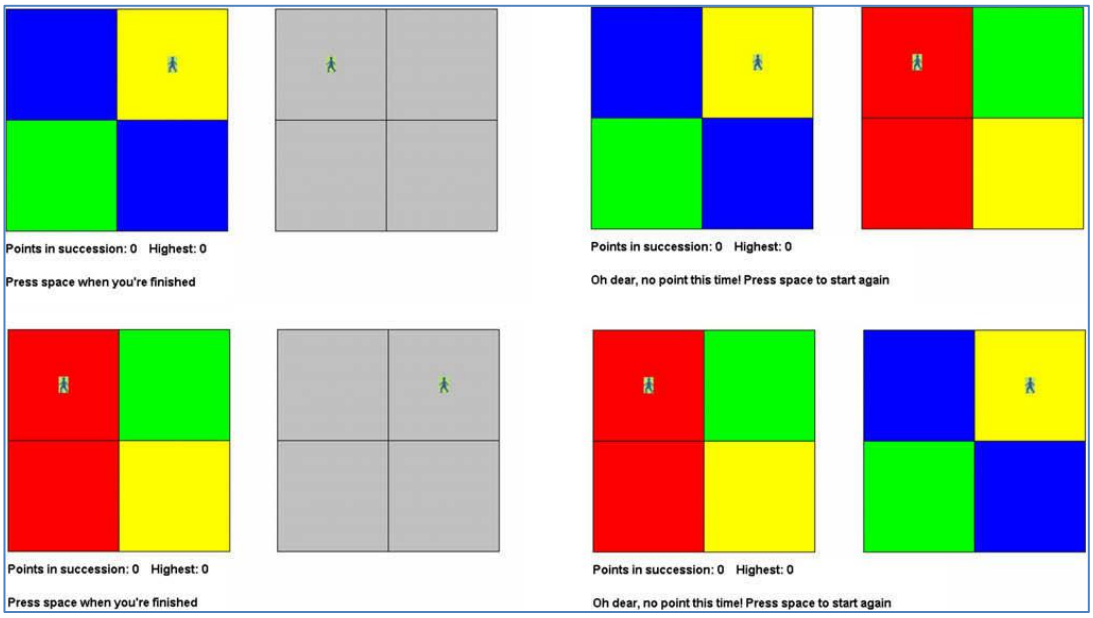


Figure 3 The experimental environment in Scott-Phillips et al. study (2009, p. 228)

Experiments:

In Scott-Phillips' 2009 study "Signaling signalhood and the emergence of communication", each participant has access to her own screen with a color vision

and the partners screen without colors. The task is moving within the 4 quadrant box, select one quadrant which has a specific color (red, blue, green, yellow) and press space to end that session, if the other participant is settled in a quadrant in her screen. Each participant can see the other's movements but not the colors of any of the quadrants that the partner is moving among. *Since the aim is to select the same color* they can see their success at the end of that session, after pressing space bar, when the color configurations of both of the participants' boxes are revealed to each other. In Figure 3, the right-hand side is the end-session view, where participants failed in both of the two rounds (selected among around 190 rounds played on average). As can be seen by comparing the round in the first row and by the round in the second row, the color combinations of quadrants change after each round. Some of the colors may be missing in some rounds but at least one common color in both of the participants' boxes is guaranteed.

Most of the couples succeeded in forming a communication system out of movements after 100 rounds of playing. They managed to negotiate and settle on a default color to select it if was available. They did this by means of non-communicative observation of selections after an average of 50 rounds. That means if they agreed upon green and if green existed in both boxes each selected (moved onto) the green quadrant and ended the round, thus scoring a point. If one considers the experiment depicted in the first row of Figure 3, one can see that if the participants agreed upon blue as default color, it would not be an option for the player on the left hand side to select it for that round. In such cases they needed to signal that "plan A is not possible" for that turn, so they developed certain unfamiliar modes of movement that signaled their signalhood (i.e., one that was distinct from the conventional movement for the sake of moving to a quadrant) and referred to a second color option which they eventually agreed upon when the default color was unavailable. Gradually they agreed on a "second color is needed to be selected" signal and a similar third and fourth signal.

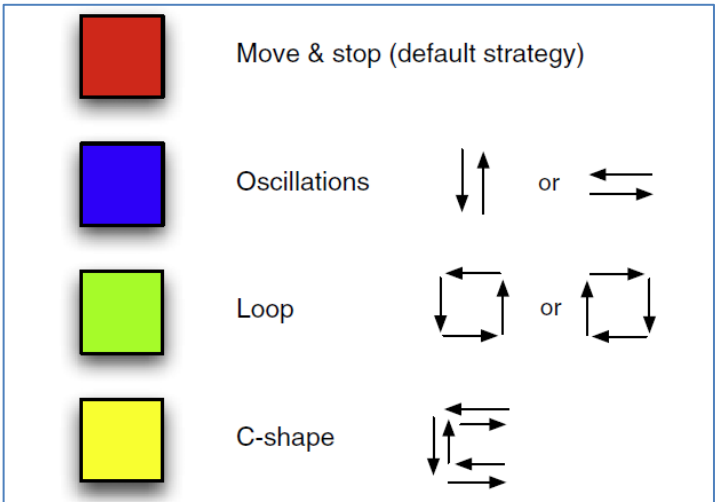


Fig. 2. A typical emergent system. In this communication system red is the default colour. If participants have a red square, they move to it and wait. If they do not have red they will signal one of the other colours by using the movements indicated. If one participant signals a colour that the other participant also has, that participant will move to the relevant square and hit space to end their turn. Otherwise, the participants will signal alternative colours until an agreement is reached. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Figure 4 Typical Emergent System (Scott-Phillips et al., 2009, p. 228)

The Embodied Communication Game proves that people achieve to signal signalhood itself, that is, they can find genuine ways to communicate even though no prior channel is provided and no instruction about communication is given. There is no experimental or task constraint that determines the way that they distinguish their signaling behaviors from other actions within the environment. The communicative system contains embodied signals.

The Bonobo Carry Activity:

The above results can be compared to observations from a different but relevant research field where animal communication sheds light on how communication or embodied communicative forms can emerge spontaneously. Captive but not human-enculturated hominids (Bonobo or *Pan Paniscus*, in particular) are very social in terms of interactions including communicative ones like greetings, threatening, grooming and playing. One of those interactions is the carry

activity of mothers. Even though the mother seems to be the primary agent in this activity, the infant may be the demanding one and the performing of the carry activity is a highly interactive task where complementary bodily coordination of the both sides is required.

> Mothers and experienced infants come together for the carry activity in a very fluid way. The transition from other activities to the carry is an almost ballistic event. Mothers often sweep up infants and move off while looking at their destination. A mother can pick up an infant without looking directly at the infant because the infant simultaneously moves its body and hands in ways that fit and take advantage of the mother's motions (Hutchins and Johnson, 2009, p. 536).

The interesting part of this interaction is about the *initiation of the carry activity*, which can be performed by the infant or the mother, such as when the infant climbs on the mother's back and the mother cooperates. But when they are outside the carry range gestures can be observed as frozen fragments of the initiation protocols of the carry activity, i.e., the infant opens its arms and poses in a frozen style, as depicted in Figure 5.

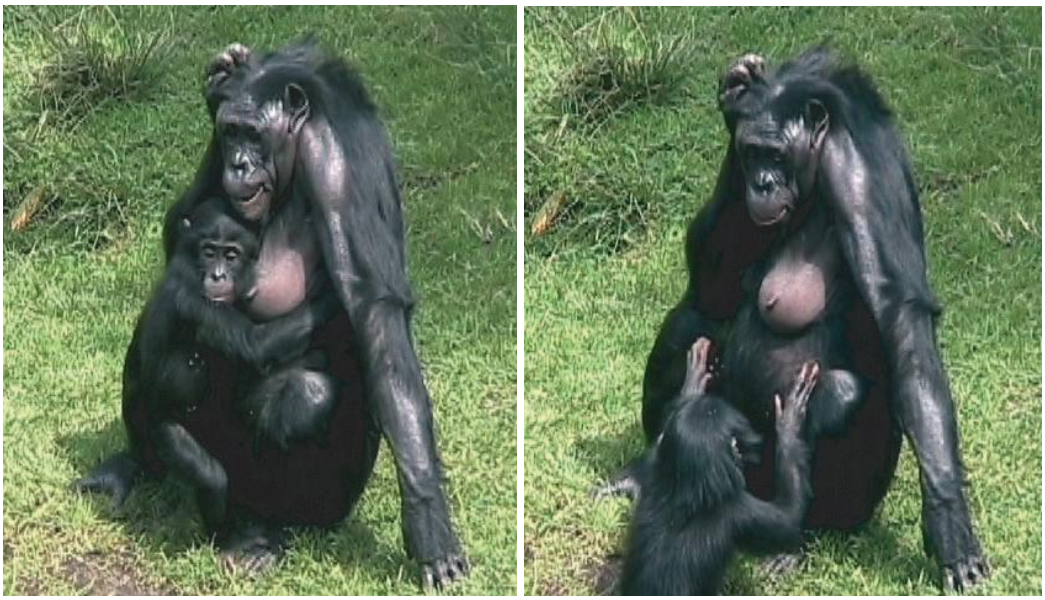


Figure 5 Gestures as frozen fragment of behaviour (Hutchins& Johnson, 2009, p. 536).

The infant Kesi clinging to her mother, Lana (A), moves away, then turns and freezes in a pose that invites being picked up (B). (Photos courtesy of Christine Johnson)

There are also "present back" gestures of the mothers to invite the distant infant that is still in her visual field. These gestures can be exemplified as the "lean back" of the infant (parts of the action series to perform the ventral carry).

All of these gestures are parts of patterned embodied complementary behaviour (rather than imitation), but later on they are used as "raw materials for meaningful gestures": "Over time, acts like lean back and reach differentiate from the normal flow of events and are displaced in space and time. They are performed not in the midst of a carry, but in its absence." (Hutchins and Johnson, 2009, p. 538)

The example of bonobos and the Embodied Communication Game (ECG) in the Scott-Phillips et al. study can be compared and contrasted. There are complex mechanisms that enable the development of coordinated, complementary action of the infant and the mother in the bonobo carry activity. The communicative action of using gestures to initiate this activity also follows a process of emergence; the development of signs as the differentiation from the general carry activity is observed. The gestures are embodied like the signals of the ECG in Scott-Phillips et al.'s experiment. However, the development and negotiation of the embodied signals in the experimental environment is subject of active reasoning of the participants. They consciously try to develop these signaling systems after realizing that they need to inform each other about the color configuration of their experimental visual field to meet the performance requirement of the experiment. This difference seems to be due to their current communication skills that they already developed by using a dedicated channel for communication, i.e., NL. The embodied signs or the communicative actions that they create *ad hoc* in the experimental setting are just similar ones they use in NL in terms of their divergence from non-communicative actions. They are consciously crafted and as complex as the task structure requires. So the existing experience and the competence of using

dedicated communication channels and tools may have facilitated the success in the experimental environment. In the sense of If the goal is observing an evolutionary or adaptive emergence of communication; similar experimental paradigms may be considered as problematic as experimental paradigms where participants are presented a readymade communication channel.

2.5. GALANTUCCI'S STUDY ON EMERGENCE OF COMMUNICATION:

The problem of the cognitive study of communication led researchers to create artificial environments, which compel human individuals to develop new ways of communication, where the interaction can be thoroughly observed and recorded. Galantucci (2005), for example, hinders participants' verbal, textual and even graphical communication and provides them only with very simple means of communication while they were trying to play a simple multi-player video-game requiring coordination and communication. His experiments formed a model or a starting point for the present study, therefore his experimental setting, research motivation and results are being presented as the latest section in this chapter.

In Galantucci's study participants interact in a virtual environment. Their task is to meet in the same room in a very coarsely designed two-dimensional maze of four rooms (see map (B) in Figure 7). They can control the movements of the agents representing themselves from a keyboard and send graphical signals to the other by drawing very simple shapes and lines to an electronic pad. The digital pad is designed to distort the graphical signals by a continuous flow (see Figure 6) in order to avoid sending obvious, clear, durable, and conventional graphical images or texts. The main questions are: (1) how will communication systems emerge and develop and (2) what are the main mappings used to represent situations, objects or places?

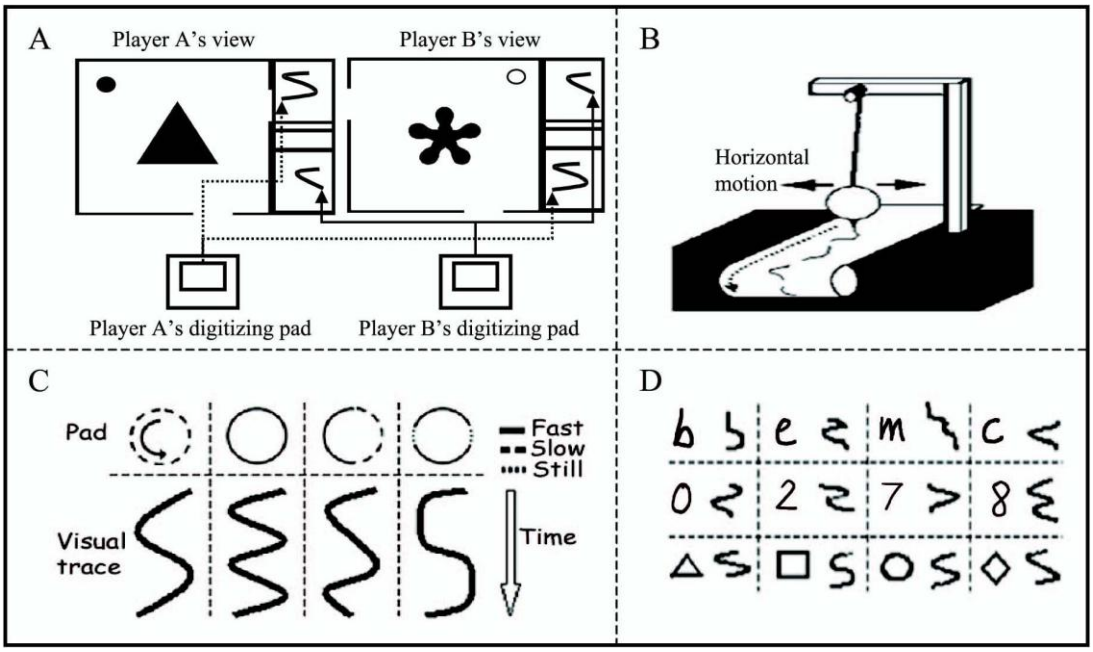


Figure 6 The communication medium in Galantucci's experiment (Galantucci, 2005, p. 741)

When one of the participants is asked to find the other or meet in any room, the subjects tend to try to communicate using the pad to let the other know where s/he is or what kind of object exists in the room that s/he is currently in (see Figure 6, Figure 7).

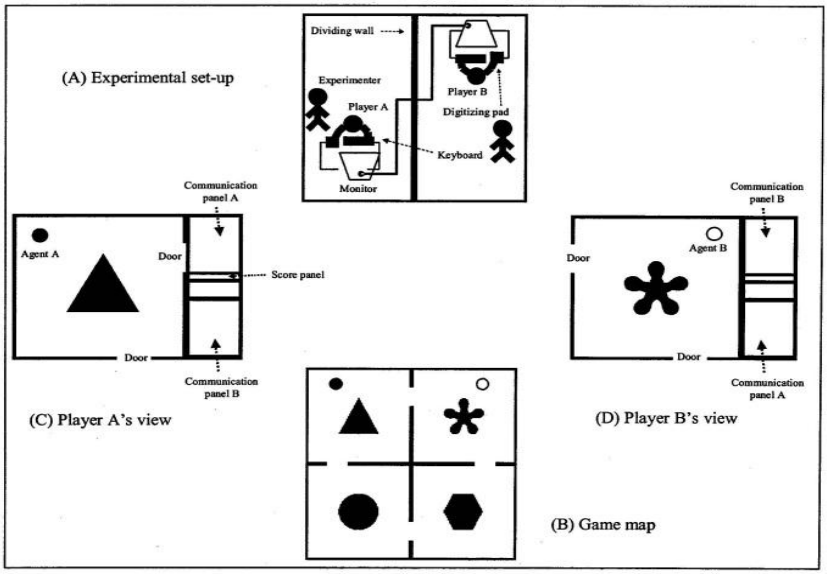


Figure 7 The maze and the rooms

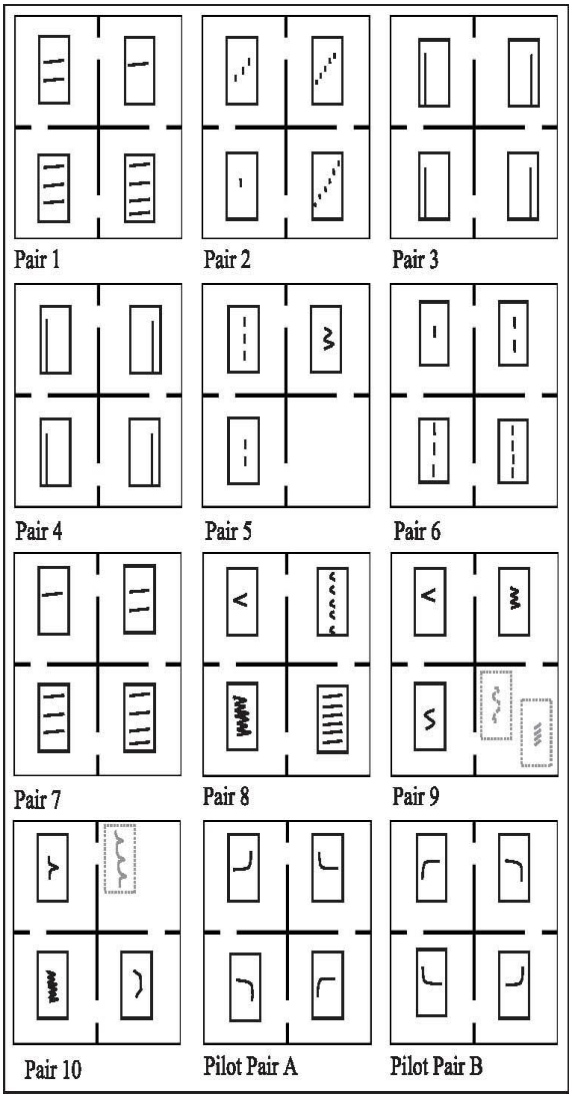


Figure 8 Some communication strategies (Galantucci, 2005, p. 740)

Participants had mainly two possible options to communicate their location. First, they can try to use signs to refer to the location of the room in the maze as did the pilot Pair A. Second, they can refer to the geometrical objects in the room that can distinguish rooms, i.e. the number of vertices (sides) of the shapes in the room, as in Pair 2. The first strategy is called “map-based” and the second is called “icon based”.

The main result of the experiment is that these real agents are able to collaborate on the task and communicate successfully. A communication system

that provides a hundred percent efficient coordination and success in each round tends to emerge relatively quickly (in 76 minutes average) and reliably during the course of the experiment. This result is not trivial since some pairs could not converge to a system at all.

The second important result was that sign systems develop parsimoniously which means that they do not contain redundant features and are specialized for the problem.

The main strategy is *Learning-by-using*: when a new signal is sent, a heuristic process starts to detect or negotiate the intended meaning; trials and errors to provide an appropriate response in consecutive sessions increase the probability of guessing the correct meaning. Similarly these trials may facilitate setting a shared meaning of the signal. This process of learning the meaning by using also helps to figure out the main strategy of creating signals.

The mapping or "naming" styles that the couples used to represent their location or the object that labels the location are as follows:

1- Map-based location information: The participants tried to communicate the location of a room (i.e., the room that they are currently in or they want to go) in the 4-room experimental environment. This also contained an iconic strategy of indicating the distinctive corner of the room that signals location. (Figure 8, Pilot pairs 1,2)

2- Numeration (Using numbers): The 4 rooms are labeled by creating numbers in a clockwise or counter-clockwise order and the numbers are represented by simple sticks. The communication pad cannot disable/distort such very simple iconic or graphic messages like straight lines or consecutive dots. (Figure 8, Pair 1 and Pair 7)

3- Iconic (using the icon of the room): Each room contains a distinctive geometrical shape (triangle, circle, hexagon, five petal flower-like star). These shapes are communicated by representing their features, namely

the number of vertices or corners. Using the number strategy, e.g., 3 dots for the triangle, one for the circle, and 6 for the hexagon are used to denote the shapes. (Figure 8, Pair 2 and Pair 7)

4- One further significant strategy is a partially complete, again a map-based strategy, used by Pair 3 and Pair 4. Here, participants draw a straight vertical line near to the left or right of the drawing pad and denote their location with only a partial representation – omitting the vertical information “up” and “down”. This can signal only the location horizontally, i.e.: “I am at one of the two right-hand side rooms of the maze” (or left-hand side rooms). But surprisingly these pairs performed best in the following sessions: they synchronized their moves with the partial location information and successfully met in the same room. Such a partially covering communication system works because it is integrated with the environmental information provided through the coordinated action.

This last result is more than relevant for the general claim (to be discussed later) that the development of a language or communication system (CS) is an integrated part of problem solving, resulting in an optimally complex solution. Hence, the communication system is not more complex than needed and is not necessarily powerful enough to represent every aspect of the objects, situations or intentions. Visual coordination completes the CS, which, at the same time, is limited by the complexity level of the environment.

The emergent communication system is embedded in the coordination of behavioral processes, and this embeddedness explains why the (relatively weak) power of the rudimentary sign system suffices to facilitate the communication (Steels, 2006a). The partial sign system that efficiently designates the location in Galantucci’s experiment presented above is a case of successful comprehension of underdetermined speaker meanings, similar to several pragmatics examples given in the relevance section. Examples like these will also occur in the present study, as

will be reported in the result section below. When people are able to integrate the communicative coordination with the behavioral coordination, this action/communicative action complex turns into the problem solving strategy for the task accomplishment.

2.6. DISCUSSION ON THE PRESENTED STUDIES ON THE EMERGENCE OF COMMUNICATION

The studies examined and presented in this chapter can be grouped as approaches using the extended/distributed cognition and the joint action paradigms, the multi-agent modeling/simulation and the adaptive/evolutionary paradigms. Of course several studies overlap in their utilization of concepts and methods of more than one paradigm or approach. These studies seem to be in dialogue with each other by adopting different paradigms' concepts and discussing each other's results in a comparative manner.

For example Nolfi (2005) combines the evolutionary/adaptive approach of Oliphant (1996) and the multi-agent modeling/simulation study of Quinn (2001) to explain how communicative action emerges from behavioral action and contributes to the joint action paradigm. This multiplicity of approaches and their interaction provide rich theoretical and methodological tools to study emergence of communication.

However, there is a shortage of attempts to unify the dedicated cognitive capacities regarding comprehension and production of communicative behaviours in these approaches or paradigms. Interestingly, the pragmatic/linguistic study of communication by Sperber and Wilson (Wilson D. & Sperber, 2002), the Relevance Theory, attempts to offer a cognitive architecture account of cognition and communication. This is achieved by presenting a dedicated sub-module of comprehension intention detection and two working principles, namely *the*

cognitive principle of relevance and *the communicative principle of relevance (optimal relevance)* (Sperber & Wilson, 1995).

Unfortunately the above-mentioned interaction and discussion between the joint action, evolutionary/adaptive and multi-agent/artificial-agent approaches does not include relevance theory, neither incorporates the theoretical concepts or approach of relevance theory, nor discusses or refutes it. The present study attempts to integrate relevance theory into the emergence of communication literature.

CHAPTER 3

3. METHODS AND EXPERIMENTS

In this study, the main idea is to create environments where individuals will have to create a communication system from a severely limited number of available symbols and to study this system's function for and relation with, the cognitive tasks that the participants will perform collaboratively.

The initial, general research questions are as follows:

1- Are individuals capable and if so, to what extent are they capable of creating a new/alternative communication system under specific linguistic and environmental conditions?

2- Are there arbitrary variations among possible communication systems that are expected to emerge or are there resilient features and trends that materialize reliably across these communication systems?

3- During the emergence of a novel but yet preliminary communication system, how do individuals reach a convention and align their communicative actions in terms of symbols or signs?

4- How is the emerging communication system related to the task at hand, the environmental constraints or features, namely to reach a certain goal in the participants' joint actions? In other words, to what degree can these communication systems be considered as an adaptive system that is inherently integrated with/determined by the requirement of a general set of adaptive skills that are constrained by the requirements and necessities of the tasks and environment?

These research questions have been addressed in a series of pilot experiments and in the main experiments of the study that explore various parameters considered important for establishing a novel communicative system. These parameters will be described after the two different kinds of pilot studies and their preliminary results in the following sections. The aim of these pilot studies was to test several platforms or environments that might be suitable for the main study. Conclusions and observations that are relevant for the design of the main experiments were drawn from these pilot studies.

3.1. PILOT EXPERIMENT 1: GOOGLE EARTH EXPERIMENT

In a first pilot study, two participants played a "location finding game" on "Google Earth" software. The task was to find a very small and unnoticeable street within a huge city (Istanbul). The first one who finds the target has to help the second participant. They communicated textually by an instant messaging software and they were only allowed to use the following symbols as "words": q, w, e, r, t, y, u, o, p, a, s, d, →, ? . They were informed that ***these letters should not be combined to create meaningful NL words***. The signs "→"and "?" were given as additional words to facilitate the initiation of speech acts, if participants prefer to. ***Usage of proper location names was allowed***, that means the users were allowed to type street names like "Sinanpaşa Cd." etc. Both participants (preferably in different rooms) could see each other's screen (her/his focus point and movement on the map) for 15 seconds in every 3 minutes, by means of graphical desktop sharing software (VNC). This screen sharing was essential and the amount of time was a result of fine tuning after a few initial trials of the pilot study. If the participants were allowed to see where the other was on the map all the time, they would coordinate their searches together with their partner or just follow where s/he was looking, without bothering themselves with trying to communicate, in fact without trying to develop a set of communication conventions from scratch. But limiting screen sharing provided the need of communication while they were doing their

own searches. A limited, short period of screen sharing provided joint attention and visual cues for coordination. After the experiment, each participant was interviewed about the intentions/meanings behind each line of his/her textual communication. Also each participant's interpretation of the other participant's message was examined in this post-experiment interview. Since the participants were planned to be taken into the next turns of the experiment with similar tasks they were not allowed to talk about the experiment before and after the pilot study. For the same reason, the post-experiment interview was carried out in the absence of the other partner.

Sample from the pilot experiment 1 data (only blue characters are the actual communication, the red characters are the interpretation of participants and notes):

==========

Date	Hour	Src. Rec.	Message
27.03.2009	01:46:39	p1 --> p2	bahceli evler FERiT SELiMPAsA CADDESi SiTE SOKAK (instruction)
27.03.2009	01:47:01	p1 --> p2	qwertyuopasd -> ? (instruction for available signals)
27.03.2009	01:52:07	p1 --> p2	? (are you there)
27.03.2009	01:52:21	p2 --> p1	s (random response)
27.03.2009	01:52:26	p2 --> p1	? (i don't understand)
27.03.2009	01:55:29	p2 --> p1	s->a (not recovered)
27.03.2009	01:55:59	p1 --> p2	b site sk (a different but the wrong site sokak (street) which is apparent in the map is detected in bahcelievler province immediatle by P1)
27.03.2009	01:56:17	p2 --> p1	t (and viewed by P2; t means tamam (alright) for P2)
27.03.2009	01:58:22	p1 --> p2	site dk (dk is a typo for sk sokak)
27.03.2009	01:58:24	p1 --> p2	sk
27.03.2009	01:58:27	p1 --> p2	site sk (correction)
27.03.2009	01:59:01	p1 --> p2	? (P1 asks for confirmation)
27.03.2009	01:59:31	p1 --> p2	b site sk (i am close to "site sokak" now on the map, b means ben) (P2 receives b as bigger)
27.03.2009	01:59:53	p1 --> p2	ferit selimpasa cd.? (but where is ferit selimpasa cd. (street) Do you see it?)
27.03.2009	02:00:55	p2 --> p1	b- >ferit selim pasa (do you mean "ferit selim pasa" by "b")
27.03.2009	02:04:37	p2 --> p1	e bagclar (this is bagcilar, e = eee? "Skip here")
27.03.2009	02:05:22	p1 --> p2	site sk bagcilar? (is your "site sk" at bagcilar)
27.03.2009	02:05:50	p1 --> p2	ferit selim pasa bagcilar? (do you see ferit selim pasa at bagcilar)
27.03.2009	02:06:25	p2 --> p1	e (negative)
27.03.2009	02:08:37	p1 --> p2	site sk? (still did not get the affirmation about P2's location at site sokak)

27.03.2009 02:08:39 p1 --> p2 e? (yes??)
27.03.2009 02:08:47 p2 --> p1 t s (t tamam s not recovered)
27.03.2009 02:09:02 p1 --> p2 ferit selim p caddesi ? (do you see fsp)

Src: source, Rec: receiver; p1: participant 1, p2: participant 2. Transcription is 8 pages in total.

The pilot study was carried out once with two different couples and twice (a second turn of the experiment with a similar task) with another couple. The initial observations were that individuals tried to project their most common NL words and speech acts onto this raw material of communication. First, they exploited the discursive properties of textual communication, location and order of proper names in a message to communicate what they see on the map at some moment. For example if a participant got guidance from the other and wanted to be sure that a street ("Site sok.") was at the east (right) of the other ("Ferit paşa sok."), the participant wrote "Ferit paşa sok. Site sok. ?"

The affirmative answer may be "t" or "y" or "o" ("tamam" (alright), "yes", "ok").

As it can be seen from the sample, they tried to assign some letters to NL words (like "t" to "tamam").

The second point is that when a verbal protocol is used to understand both participants' intentions for their communicative actions (in the post-experiment interview), we can see that they mostly misinterpret each other (in particular, in the beginning), however, if the "will" to communicate was maintained, they managed to establish tiny but reliable protocols, usually accidentally. So the factors of "willingness to communicate" and the confidence in each other's reliability in sending consistent messages were some other crucial psychological factors. This was one of the results and observations of Galantucci's 2005 study as well, which he called it "learning-by-using" (p. 745). The participant may have misunderstood or did not understand at all but they had a confidence that the same signal would be

used in a similar context and after some trial and errors the context would be revealed. This confidence is critical for the triggering of initial conventions and settling on the strategy to exploit existing/shared communication skills.

The most important outcome of this pilot experiment was the realization of the fact that this study may have been a study of emergence of communication between individuals who already have acquired a perfect system, namely NL. Thus, also the present study might not investigate the invention of completely novel communication systems but rather investigate the modalities of new but re-emerging systems, on the background of the already existing one and under the contextual constraints/requirements of the experimental environment and task.

3.2. PILOT EXPERIMENT 2: ACTIVEWORLDS ENVIRONMENT

The pilot experiments carried out in the Google Earth environment were problematic in terms of task simplicity and the moves in the environment required for the task accomplishment, not being real-world like. The participants were virtually flying over the city, descending down (i.e. zooming in) to see a small location and then ascending (zooming out), etc. The only fruitful feature of the Google Earth experiment was allowing participants to use proper names of locations in the city. This facilitated the communication (no need to reach a convention on a lexicon for location names) and participants just exploited given lexical items (letters) for words of speech acts and for function words (like pronoun, verb, prepositions etc.).

ActiveWorlds on the other hand, is an internet based, interactive, 3D, multi-agent environment software with features allowing the researcher to create and customize his/her own site to carry out experiments. The latest version used during the experiments was ActiveWorlds Browser 6.0 Enterprise, (see: www.activeworlds.com). In this virtual environment participants walk on the ground and wander in the area, have a first person view of the site (Figure 9) and can interact with some objects. Crucially, agents can textually communicate during

that time. Using additional software, it is also possible to let participants communicate with auditory signs.

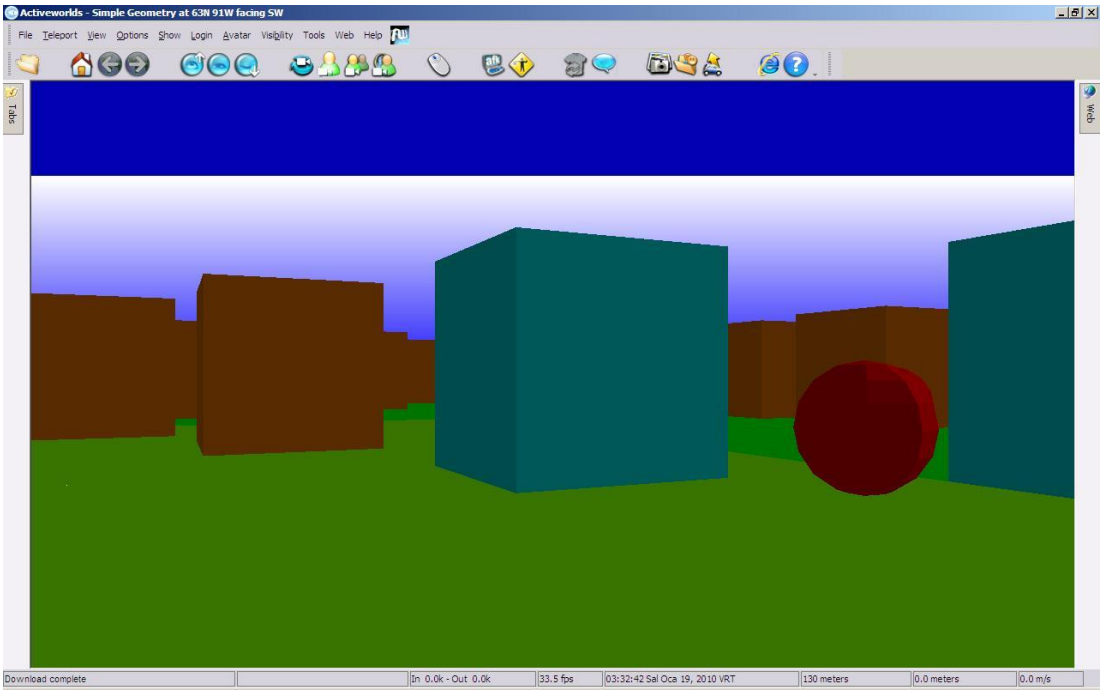


Figure 9 First-person view of the ActiveWorlds environment by the participating agent in the experiment

Using this environment a small maze of 3D objects (Figure 10) was created. As it can be seen in Figure 9 participants can see the others bodies, “avatars” of the other, if there is no visual obstacle in between, like in the real world. In this situation and/or when there is no visual contact, they can textually communicate with the built-in chat utility or a third party instant messaging program.

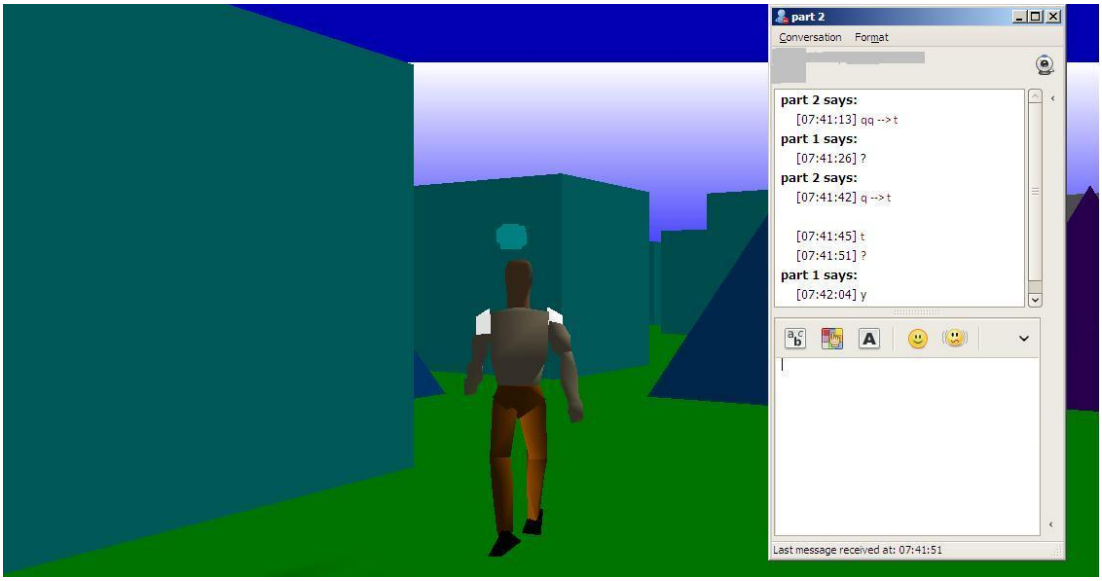


Figure 10 Second-person view of the participating agent

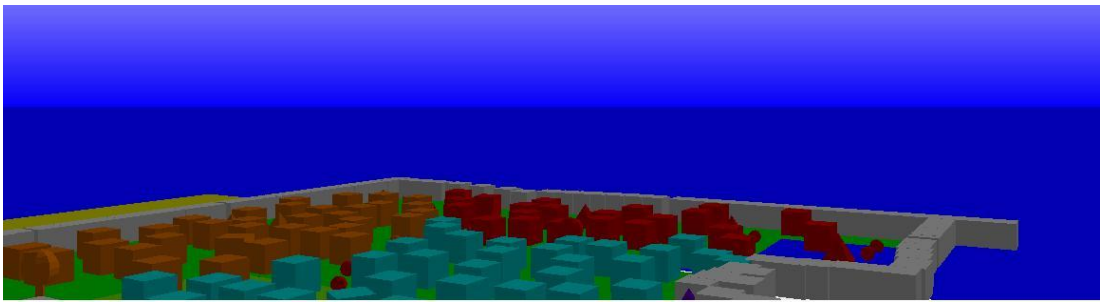


Figure 11 General view of the 3D maze (note that participants cannot view the site in this panoramic style)

The task, constraints on the textual communication and the inventory of lexical items (q, w, e, r, t, y, u, ı, o, p, ?, →) were similar to the previous pilot study. The task was to find an object and after that, find and help the other participant who may be looking elsewhere and lead her to the target. There was no actual time limit. Participants were informed about the way and rules of communication but not asked to communicate textually, unless they needed and wanted to do.

Results of the first pilot studies in the ActiveWorlds environment showed that given the same constraints with the previous pilot experiment, participants

tended to communicate less in comparison with the earlier experiment. The reason seems obvious: here there is no lexicon available for the main items in the scene. The main items were location names and these proper names were ready for utilization in the Google Earth experiment. The participants tended to communicate in terms of speech act words such as “eee?” “oq” “qoo” to direct or call each other; however, usually this did not help much. Users tended to find each other, and coordinate the search of the target by keeping visual contact. This provided joint attention, at least in terms of sharing the same portion of the 3D maze and searching together. Nevertheless, even in the case of scarcity of communication in the ActiveWorlds environment, it was observed that, participants managed to build a few communicative conventions which helped them to collaborate, if they spent time in the environment and kept communicative actions at a certain frequency. Parameters such as the time span of the virtual experience, the psychological condition of willingness to communicate between the participants, and the emotional attitudes towards the initial attempts to communicate (which usually failed and annoyed both participants) are very decisive. Lack of patience or disbelief in the communicative success usually led to quitting the experiment or completing the task without communication. If one participant was interacting with a friend of his, it was psychologically easier to be patient and keep on trying to communicate. So the initial belief on the partners’ communicative skills appeared to be an important parameter or constraint. The significance of these emotional parameters, however, is not investigated in the present study.

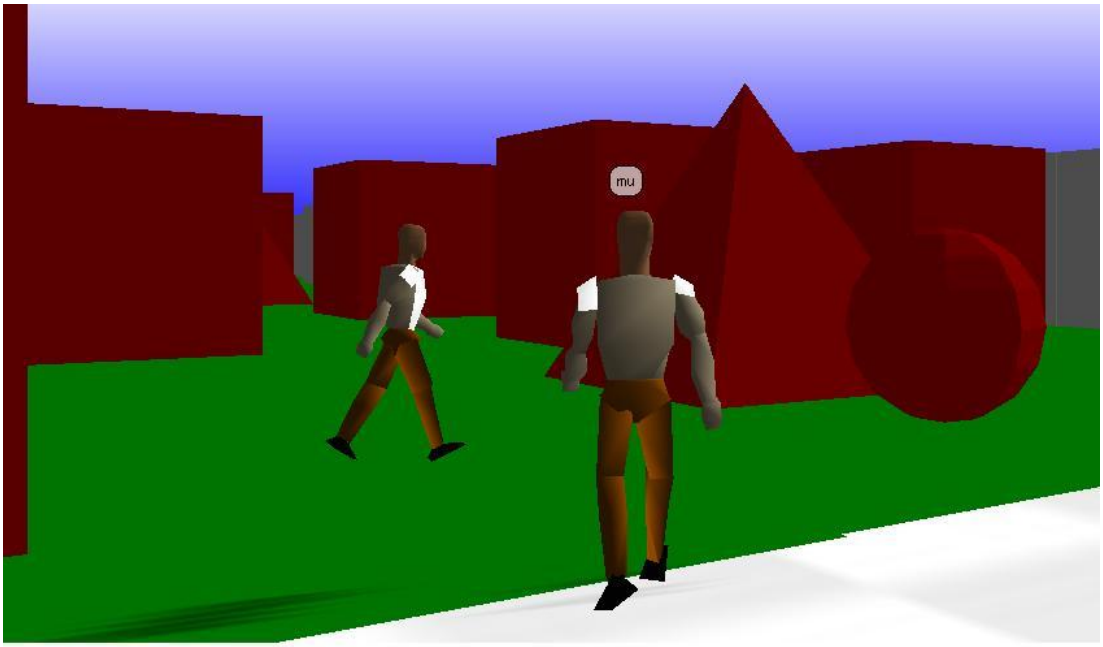


Figure 12 ActiveWorlds context of for the sample communication

Results of the Pilot Experiments in ActiveWorlds:

Without any instruction for or guidance to play a naming game, in the current ActiveWorlds environment, participants tended to communicate less in comparison with the Google Earth experiment since they had ready lexical items for locations. The advantage of ActiveWorlds is that the experimental scenario can be easily changed to facilitate or enforce communication. The reason for less communication seems to be the task itself, for finding locations in this small 3D maze can be done by individually searching the target or visually following the partner. But the ActiveWorlds environment is very suitable for changing the scenario into anything. For example, if the target is also in motion, like a third agent to be chased or caught, the need for coordination increases and this requires communication. As we can see from samples in the following sections, if such a communication system is needed, agents seem to be capable of creating it, usually exploiting the existing NL skills and projecting the existing NL conventions on the given tools of communication. The next important thing was that the main items

being developed were “colors” and “object types”. The interesting thing was that the 3D maze was designed in such a way that all objects were grouped in three regions and three colors, i.e., every object was in similar color with its neighbors. So if we assume that there is an economy principle governing the complexity of the communication system, we may expect the participants to use colors not only for naming the colors of cubes and pyramids but as region names in the maze also.

3.3. COMPARISON WITH GALANTUCCI’S EXPERIMENT

Galantucci argues that most of the studies about human communication either rely on NL or provide an established artificial language to participants designed by researchers. He states that these studies “tap into the processes that lead to emergence of communication systems only indirectly” and proposes a research environment where participants cannot rely on pre-established languages “to tap into the processes that lead to emergence of communication more directly” (Galantucci, 2005, p. 738). Such an environment requires the absence of pre-established human communication systems like writing and speaking.

Our experimental environment and initial conditions can be thought of being somewhere in the middle as compared to Galantucci and the previous studies on NL acquisition and communication. In the Google Earth experiment participants were given random letters as candidates for lexical items but no additional guidance or direction about how to use them for communication was given to them. In the ActiveWorlds environment, users had the same kind of inventory of lexical items, directions and prohibitions. The difference with Galantucci’s experiment is that, even with random, meaningless lexical items and given a general framework of textual communication, participants tend to exploit their previously existing competence of communication and the shared NL. They exploit the order of lexical items since textual communication is both linear and visual (unlike speaking where lexical items are not visual), as when they try to represent the spatial organization

of the items in the virtual environment in the message they sent etc. They try to mimic speech acts like questions and commands.

This makes the present study one that looks for the possible ways of building a communication system collaboratively by participants, who inevitably have an earlier skill of utilizing a communication system. But when we look at the discussion part of Galantucci's work, the very same kind of exploitation of general/previously acquired skills of communication is detected (p. 750, section 3.1.6.1). Participants exploited previously acquired and shared representations and features of geometric objects, icons, number systems and map-using conventions. Even in the Scott-Phillips et al. 2009 study, where there was no communication channel at all, the participants used prototypical circular or oscillatory periodic movements or C-shape movements to signal the saliency (signalhood) of their communicative action as distinct from their regular task related movements in the environment. In conclusion, when there are real human agents it is impossible to conduct a pure emergence of communication experiment that prevents exploitation of existing communicative skills and shared knowledge.

3.4. THE PARAMETERS AND THE MODIFIED AND FINALIZED DESIGN OF THE PRESENT EXPERIMENTS

In the light of the two types of pilot experiments conducted previously, the structure of the main experiments was modified as follows:

1- Participants are assigned to a variety of tasks that will instigate them to communicate. The environment actually is modified accordingly, in order to minimize continuous visual contact. This is accomplished by increasing the numbers of 3D objects which decreases the sight distance.

2- Experiments are conducted with pairs of participants; hence, the population of speakers does not consist of 3 or 15 people in the same experimental environment but of dyads). The emerging communication system will be the

specific language observed among the two. The main advantage of restricting the setting to a dyad is to limit the management and interpretation of data.

3- Each couple attends a new session of the experiment with a similar task if they manage to initiate communication even with few lexical items agreed upon. The number of sessions/ experiments is increased to eight to follow the development or "evolution" of the emergent communication system. This helped to test the hypothesis whether "individuals are capable of developing new communication systems" and the hypothesis is that "this communication system tends to get more complex under continuing interaction".

4- In order to test the hypothesis that "individuals tend to align with lexical preferences of the community or the partner of the communicative action during interaction", the patterns of comprehension and adoption of the other individuals' usage will be tracked.

3.5. GENERAL HYPOTHESIS AND RESEARCH PARAMETERS

The emergence and development of communication systems are studied from a variety of theoretical perspectives and experimental methods. These are the multi-agent systems approach, semiotic dynamics, cognitive theory of communication (relevance theory) and the evolutionary approach.

After a fine-tuning of experimental environments that facilitate the emergence of communication systems the following results are expected:

1- Humans have the cognitive ability to develop and adopt inventions/novelties in a communication system. Some level of sophistication in the system will be observed when the community of agents is exposed to the system with some persistency.

2- The level of sophistication of the emergent system is expected to be limited to the complexity of the environment and the task (i.e. problem space).

Firstly, these results can be theoretically articulated within the existing literature on human cognitive agents' capacities of collaboratively forming semiotic systems, individually managing the cognitive load of shared attention and handling the communication with other agents.

Secondly the emerged communication system can be depicted and described as not only as a communication tool, but its integration, parsimony and embedded nature within the structure of the environment and with the problem/task can be analyzed.

Thirdly, following the multi-agent systems analysis and the semiotic dynamics studies, the results can be tested and reciprocally used to evaluate results with artificial agents and real human agents observed in communicative/ interactive/social contexts. This potential use of the results, namely the comparative analysis, is, however, not in the scope of the present study.

3.6. MAIN EXPERIMENTS AND DATA PROCESSING FOR DESCRIPTIVE ANALYSIS:

The descriptive analysis of the data collected from a total of 64 experiment sessions conducted with eight distinct couples is presented. The experimental environment is created and designed by a 3D computer environment tool (called ActiveWorlds).

Eight couples' communicative activity (restricted by experimental design) is recorded, participants are interviewed after each experiment and their explanations are transcribed. This quantitative data is analyzed and demonstrated on four levels:

1- **General quantitative and temporal information** about the 8 experiments for each couple is collected and analyzed.

2- The content of the communication is presented and categorized into **lexical categories** that exist in NL.
3- A **syntactic complexity measure** is offered and used to categorize each utterance in each turn taken by a participant.
4- **Speech act categories** are used for describing the types of communicative actions that took place during the experiments.

The common properties, characteristics of the communication system developed by the two couples and corresponding quantitative data is presented and discussed. The dissimilarities and idiosyncratic issues are explained.

3.6.1. EXPERIMENTAL DESIGN

The allowed symbols are ? and → (only the right arrow symbol) and the letters are q, w, e, r, t, y, u, I, o, p.

Uppercase is allowed. Combination of the letters into words is allowed. An information sheet about the rules and the tasks of the experiments (in Turkish) is given to the participants (see Appendix A).

The first six of a total of eight experimental sessions were conducted usually within a week, where some were successive, (like the first two sessions completed within 2 hours) the last four experiments in a second or third day (for example; 2 sessions on Monday, 2 sessions on Thursday and 2 sessions on Saturday). Each session takes nearly one hour, preparation of the participants and the computer environment (around 10 minutes), experiment session (around 10 minutes) and separate post-experimental interviews with each of the participants (20 minutes). All 8 experimental sessions with the 8 couples required at least 192 hours involving at least 3 persons (two participants and minimum one experimenter)

3.6.1.1. TASKS/TARGET OBJECTS:

The tasks comprise finding a unique (in terms of shape and color combination) object in the environment and leading the other participant to the object.

After the first 3 sessions the number of targets given to the couple is increased to two, one for each.

After the first 6 sessions, a nearly two-month break is given before the 7th and 8th experimental sessions are conducted with the same couples. The effect of this break (if observed) will be used to check the function of long-term memory capacities of the dyads and resilience of the developed communication system.

The participants are expected to complete the tasks collaboratively and in coordination. The task usually requires collaboration and communication by its very nature. Both of the participants are requested to find the same unique object in the area but the explicit task is given to only one of them. So the content of the task requires to be communicated. In the first three experiments (1-2-3) there is single task given to one of the participants. In the second group of the subsequent five experiments (4-5-6-7-8), there are two targets to be found, but each participant receives only one of them. All tasks, including English translations for the 6+2 experiments, are given in Appendix B.

A sample from the task list of the 8 experiments is as follows: P1 (participant 1) P2 (participant 2).

Experiment 1: P1: Find the blue sphere and take the other participant to it. (Mavi küreyi bul, digger katılımcıyı kürenin yanına götür)

P2: The task is given to the other participant. (Görevi diğer katılımcıya söyledik)

Experiment 4: P1:. Find the green sphere and take the other participant to it. (Yeşil küreyi bul; digger katılımcıyı kürenin yanına götür)

P2: Find the green cube and take the other participant to it. (Yeşil küpü bul; diğer katılımcıyı küpün yanına götür)

Table 1 The 8 experiments' task list

	PARTICIPANT 1	PARTICIPANT 2	
EXP1	**Blue Sphere**	--	SINGLE TARGET
EXP2	**Green Cube**	--	SINGLE TARGET
EXP3	--	**Yellow Pyramid**	SINGLE TARGET
EXP4	**Green Sphere**	**Green Cube**	DOUBLE TARGET
EXP5	**Green Pyramid**	**Brown Sphere**	DOUBLE TARGET
EXP6	**Purple Cube**	**Purple Pyramid**	DOUBLE TARGET
TWO MONTHS GAP			
EXP7	**Purple Sphere**	**Green Cube**	DOUBLE TARGET
EXP8	**Green Sphere**	**Green Pyramid**	DOUBLE TARGET

For each experimental session the given targets are unique in the environment. To facilitate the re-use of already created lexical items, similar targets are given (blue sphere, green sphere), sometimes exactly the same object and color combination is given, in different sessions. However, the target is not in the same location in the case of recurring targets. To avoid the effect of familiarity with the recurrent targets, the experimental environment is minimally modified between sessions. The general structure is preserved, only the colors of the target objects are manipulated for that session. For example in Experimental Session 2 (Exp2) and Exp4 there are green cubes as targets. Before Exp4, the green cube's (the target of Exp2) color is changed into red and another cube in a different location of the maze is modified into a green one.

3.6.1.2. PARTICIPANTS:

A total of 22 subjects (11 couples) participated voluntarily in the main experiment phase. One of the couples (dyads) failed to initiate a communication solution and quitted after the second experimental session due to frustration and reduced motivation. One couple's data is not used since one of the participants was not available in the second phase of the 6+2 experiments. The last couple's data is not used for analysis since the experiments could not be completed after the data analysis was done for the remaining eight couples. 14 out of the remaining 16 participants (11 female, 5 male) were METU students. All are either undergraduate or graduate students. These volunteers were paired as couples mostly depending on the match of schedules. Some members of the couples knew each other but they were instructed not to communicate about the experiments between sessions.

3.6.1.3 POST EXPERIMENT INTERVIEW:

After the experimental task is completed successfully or participants fail to communicate, fail to find the target or find each other and give up (within 600 seconds generally) the experiment session ends. If the participants are motivated to accomplish the task they are allowed to play for a longer period.

Each participant (alone) is taken to a post experiment interview.

In this interview each of the participants are asked about their communicative intentions in their turn of the conversation (i.e., what they have meant by what they have said), and their comprehension of the utterances in the turns taken by the other participant. There is a post experiment interview session (two separate interviews with both participants) in between every experimental session.

A sample of the data gathered by these interviews is a follows:

Table 2 Sample from post experiment interview records

1	2	3	4	5	6	7	8
LOG	*partcp1 Interpretation*	*partcp2 Interpretation*	*partcp1 Intention*	*partcp2 Intention*	*context*	*S/F*	*Speech Acts*
(4:53:46 PM) partcp1: yepıy prwit		yeşil piramit	yeşil piramit (bulalım)		no visual contact	Success	Commissive
(4:53:55 PM) partcp1: put		bul	bul			Success	Commissive
(4:54:00 PM) partcp2: qeywe req qure put	kahve renk küre bul			kahve renk küre bul		Success	Directive
(4:54:16 PM) partcp1: oq		ok	ok (for turn1 of p2)			Success	Commissive
(4:54:16 PM) partcp2: ırq yepıy prwit oq	ilk yeşil piramiti (bulalım) ok			ilk yeşil piramiti (bulalım) ok	no visual contact	Success	Commissive

The 3rd and 4th columns are the speaker meaning and the comprehension of utterances of P1 in her turns. The 2nd and 5th columns are those of P2 utterances, respectively.

The 7th column is about the **turn success**, which is defined as follows: "A turn is considered to be successful if the intention and the comprehension of the utterance matches".

The 6th "Context" column contains notes about the context at the time of that turn; visual contact of the participants (when they see, find and chase each other), the intermediate stages of task accomplishment (finding one of the two targets), significant situations (misunderstandings, misrepresentation of target, participant errors, loosing visual contact, the followed participant etc.) that affects the course of the events.

3.6.2. DATA RECORDING AND COLLECTION:

A sample of communication is provided below from one of the couples whose data is used in the general descriptive analysis. The logs of instant messaging software are formatted into a spreadsheet table, as shown in Table 3:

Table 3 Sample of processed communication log

turn #	Time	Participant	Message	gap (sec.)	S/F	Chracter Length	#lexical items
22	3:04:44	partcp1:	oy oy		fail	5	2
23	03:04:59	partcp2:	ret qup	00:00:15	success	7	2
24	03:05:50	partcp2:	Qo to ret qup	00:00:51	success	13	4
25	03:08:01	partcp2:	qo wıt wo	00:02:11	fail	9	3
26	03:08:10	partcp2:	qrı wo	00:00:09	success	6	2
27	03:08:25	partcp1:	oq	00:00:15	success	2	1

3.7. DATA ANALYSIS:

After the completion of all 8 experiments for each couple, the data collected is filtered and transcribed into several form tables of which different data can be extracted. The following section (3.7.1) about the quantitative parameters, presents the data analysis processes in more detail than the following similar sections about the other levels of analysis. The presented steps of data processing and analysis may be considered as model for the other remaining levels of parameters, lexical categories, syntactic complexity and speech act analysis.

3.7.1. QUANTITATIVE AND TEMPORAL ANALYSIS:

The data is compiled into an 8 worksheets spreadsheet where the quantitative and temporal parameters are measured, counted and calculated partially in an automated fashion. For each experimental session, the parameters such as temporal gaps [interval] between turn taking of the participants (in seconds), total experiment session duration, number of characters used, lexical items etc. are demonstrated in terms of totals and/or averages. All these parameters are as follows:

Table 4 List of Quantitative Parameters for a Couple's session data

1. Turn Number	13. Total time (sec.)
2. Time	14. Task Result: Success/Fail
3. Participant	15. Number of P1 Turn takings
4. Message	16. Number of P2 Turn takings
5. Gap [interval] btw. turns (sec.)	17. Total Number of Turns
6. Turn Success / Failure	18. Number of Successful Turns
7. Character Length (for each turn)	19. Number of Failed Turns
8. Number of lexical items	20. P1 ∑ lex. İtems
9. Complexity (of sentence)	21. P2 ∑ lex. İtems
10. P1 ∑ character length	22. ∑ new lexical items in each exp.
11. P2 ∑ character length	23. ∑ lexical items (type)
12. ∑ character length P1P2	24. ∑ lexical items (token)

A sample Quantitative and Temporal Analysis data file document (eight sheets for eight sessions) is available in Table 5

Using the data in the format presented in Table 5**Error! Reference source ot found.** (in the next page), a summary data file for each couples' performance in the 8 experiments (in a single sheet) is created in terms of **totals** and **averages** for each experimental session (see Table 6).

Table 5 A sample Quantitative and Temporal Analysis for Couple 8 Experiment 7

Message	gap (sec.)	S/F	Chracter Length	# lexical items	complexity	P1 ∑ character length	P2 ∑ character length	∑ character length P1P2	P1 ∑ lex. İtems	P2 ∑ lex. İtems	∑ lexical new items	∑ lexical items (type)	∑ lexical items (token)	Total time:	Task Result	
u trqt?		Success	7	2	3	48	76	124	14	23	0	15	37	00:12:22	Success	
yee	00:00:05	Success	3	1	1									742 sec		
ı too	00:00:08	Success	5	2	2										Turn	Result
qree qup	00:00:05	Success	8	2	2										S	19
purp qure	00:00:16	Success	9	2	2										F	2
qo qo	00:00:48	Success	5	2	2										Total Turns	21
wer re you	00:00:47	Success	10	3	3		vowels p1	vowels p2	vowels p1 p2							
trqet	00:00:31	Success	5	1	1		16	26	42						p1 turns	9
yee	00:00:01	Success	3	1	1		consonants p1	consonants p2	consonants p1 p2						p2 turns	12
yee	00:00:09	Success	3	1	1											
go go purp qup	00:00:10	Success	14	4	3		23	36	59							
->	00:00:13	Success	2	1	1		v/c p1	v/c p2	v/c p1 p2							
you qo	00:00:03	Success	6	2	3		0,695652174	0,722222222	0,711864407							
you qooo	00:05:09	fail	8	2	3											
wer r you	00:01:58	Success	9	3	3											
qo->ı	00:00:09	Success	6	2	3		Type 1	Type 2	Type 3							
qo qo	00:00:11	Success	5	2	2		8	6	7							
trqt	00:01:17	Success	4	1	1											
purp?	00:00:10	fail	5	1	2											
yeee	00:00:06	Success	4	1	1											
yee	00:00:06	Success	3	1	1											
			124	37	41											

Table 6 A summary of quantitative parameters for the 8 experiments of Couple 8

Couple 8 (O_D)	S/F	Time (sec)	∑ # Turns	Turn Success Ratio	∑ # lexical items	∑ Chracter Length	Average Lexical İtem length	complexity point	# p1 Turns	# p2Turns	p1 / p2 turn ratio	P1 ∑ character lenght	P2 ∑ character lenght	P1 ∑ lex. İtems	P2 ∑ lex. İtems	time normalized
exp 1	F	671	15,00	0,27	17	30	1,76	20	9,00	6,00	1,50	14	16	9	8	67,1
exp 2	S	563	17,00	1,00	26	89	3,42	29	8,00	9,00	0,89	46	43	15	11	56,3
exp 3	S	237	15,00	1,00	23	94	4,09	28	5,00	10,00	0,50	37	57	8	15	23,7
exp 4	S	689	36,00	0,94	88	303	3,44	84	20,00	16,00	1,25	158	145	45	43	68,9
exp 5	S	429	23,00	1,00	41	147	3,59	49	14,00	9,00	1,56	95	52	25	16	42,9
exp 6	S	341	32,00	1,00	57	188	3,30	58	13,00	19,00	0,68	85	103	25	32	34,1
exp7	S	742	21,00	0,90	37	124	3,35	41	9,00	12,00	0,75	48	76	14	23	74,2
exp8	S	305	15,00	1,00	24	83	3,46	27	8,00	7,00	1,14	46	37	13	11	30,5
									10,75	11,00	1,03					49,7125

couple 8 (O_D)

couple 8 (O_D)

	∑ lexical new items (x)	∑ lexical items (type) (y)	∑ lexical items (token) (z)	x/y	y/z	x/z
exp1	5	5	17	1,00	0,29	0,29
exp2	10	15	26	0,67	0,58	0,38
exp3	4	15	23	0,27	0,65	0,17
exp4	7	22	88	0,32	0,25	0,08
exp5	3	17	41	0,18	0,41	0,07
exp6	0	15	57	0,00	0,26	0,00
exp7	0	15	37	0,00	0,41	0,00
exp8	0	13	24	0,00	0,54	0,00

	∑ Complexity Point	Number of Type 1	Number of Type 2	Number of Type 3
exp1	20	11	3	1
exp2	29	9	4	4
exp3	28	5	7	3
exp4	84	6	12	18
exp5	49	7	6	10
exp6	58	13	12	7
exp7	41	8	6	7
exp8	27	6	6	3

Using this summary data file, the variation of some of the quantitative parameters across the 8 experiments are plotted to visualize and explore the change and efficiency of the communication system for the selected couple.

The summary data files of the eight couples are unified into a final 8 experiment by 8 couples spreadsheet document where all information about individual couples and their 8 experimental sessions are contained (in averages and totals) as separate worksheets (see Table 5). The tables of each parameter and their relative trends throughout the 8 experiments can be plotted and explored as in the following example:

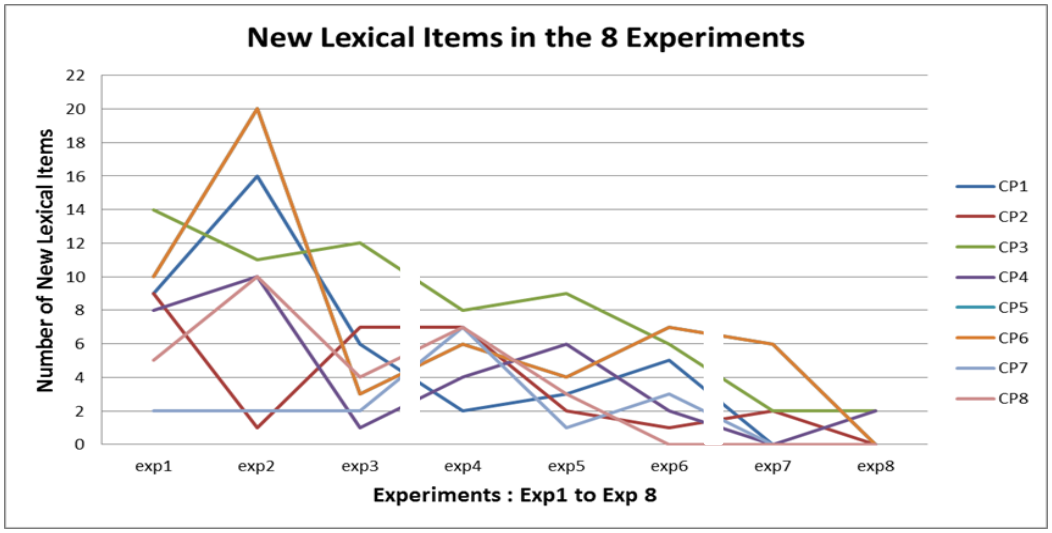


Figure 13 A sample graph of a quantitative parameter: New Lexical Items in 8 Sessions. The blue and green transparent columns represent the two interventions after Experiment 3 and Experiment 6

In the above graph, the number of **new lexical items** (which are the words used for the first time in a session, successfully added to the lexicon) is plotted across the eight experiments. Each couple is plotted as separate series (CP1, CP2,..., CP8).

Remark: The unlucky coincidence of having an equal number of subjects and experiments (8 couples and 8 experimental sessions) may cause confusion in the interpretation of the tables and the graphs. For every section of the present study, for all of the graphs, the horizontal axis contains the time variable, i.e. the successive eight experiments. Using similar graphs for other quantitative parameters, initial exploratory observations can be made about the trends and variance regarding to that parameter. For example, using the graph in Figure 13, we can say that the number of new lexical items created in the later sessions tends to decrease.

3.7.2. PARAMETRIC STATISTICS: REPEATED MEASURES ANOVA

The data explored and used for the graphs of the quantitative parameters is also utilized for the statistical test of the variance. Repeated measures ANOVA is used for the prototypical case of deciding whether a change of a dependent parameter occurred for a group of subjects that are exposed to same kind of treatment, exercise, training or external effect repeatedly at consecutive times.

Repeated measures ANOVA is defined as the Analysis of Variance "in which subjects are measured more than once to determine whether a statistically significant change has occurred" (Vogt, 1999, 336). Repeated measures ANOVA is "an extension of Paired T-Tests. Like T-Tests, a repeated measure ANOVA gives us the statistic tools to determine whether or not change has occurred over time. Repeated Measures ANOVA compares the average score at multiple time periods for a single group of subjects." (Schwab, 2007, 2)

In our case, the dependent variables are any of the parameters defined for the four levels of analysis. For the quantitative ones, our current example, "**the number of new lexical items created in an experimental session**" is the dependent variable that characterizes the communication system and it changes due to the continuous change of the dyad's communication during the eight experimental

sessions. So the subjects are not the 16 individuals, but the 8 couples or the 8 communication systems that are developing as the collaborative products of these couples. All of the parameters presented and investigated are describing features of the communication systems. The ones that contain the individual performances are used as the ratios or totals of both of the participants. In this way, again, they are considered as the features of the mutual communication system. For example, the turn taking ratio of P1 (participant 1) and P2 (participant 2), p1/p2 is in fact a measure of the symmetric or asymmetric participations in the communication system, and this ratio will be hypothesized to converge to 1 in the later experimental sessions, as a characterization of efficient use of the communication system "by the dyad" in terms of symmetric participation in coordination and information sharing.

Thus, the subjects are the couples or communication systems created by them, respectively, whereas the independent variable is the time, which consists of eight time levels corresponding to eight experimental sessions. For the dependent variable **"the number of new lexical items created in an experimental session"** we can make the assumption that this parameter is both characterizing a feature of the emerging communication system and the dyads' performance (but not the individual performances). Using the Repeated Measures Anova test the change of **the number of new lexical items** across the eight experimental sessions was found as significant and reported with the corresponding F and p values as follows: ($F(7, 49)=8.36$, $p<.001$, $\eta p2=0.544$).

Table 7 The probability of F-statistic for Repeated Measures Anova

Tests of Within-Subjects Effects

Measure: Number_of_new_lexical_items

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
new_lex_item_8	Sphericity Assumed	517,984	7	73,998	8,357	,000	,544
	Greenhouse-Geisser	517,984	2,491	207,959	8,357	,002	,544
	Huynh-Feldt	517,984	3,976	130,294	8,357	,000	,544
	Lower-bound	517,984	1,000	517,984	8,357	,023	,544
Error(new_lex_item_8)	Sphericity Assumed	433,891	49	8,855			
	Greenhouse-Geisser	433,891	17,436	24,885			
	Huynh-Feldt	433,891	27,829	15,592			
	Lower-bound	433,891	7,000	61,984			

To use these values for the rejection of the null hypothesis that “there was no change in the number of new lexical items across eight experiments”, there is a requirement of satisfying the statistical assumption of sphericity, important for repeated-measures ANOVAs.

> “When it is violated, F values will be positively biased. Researchers adjust for this bias by raising the critical value of F needed to attain statistical significance. Mauchly’s test for sphericity is the most common way to see whether the assumption has been met. (Vogt, 1999).

For the **number of new lexical items** Mauchly’s test for sphericity is satisfied (see Table 8). If it was not the case the F and p values calculated under the Greenhouse-Geisser correction should be used and reported.

Table 8 Mauchly’s test for sphericity is satisfied p > 0.05

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
time_new_lex_1_8	,002	27,106	27	,660	,356	,568	,143

Similar tests for the other parameters out of the 24 quantitative parameters are possible. However, a selective and exploratory approach is adopted to create and test the hypotheses. Construction of hypothesis is based on initial observations during the pilots, but during the data analysis phase there were unexpected significant trends noticed and tested, which were not formulized into hypotheses before then.

3.7.3. NON-PARAMETRIC STATISTICS: FRIEDMAN'S TEST

One of the assumptions for the Repeated Measures ANOVA is that the data should be normally distributed at each time level. The test used to check this is the Saphiro-Wilks test for normality which is stated as appropriate for small samples. (de Smith, 2011) (http://www.statsref.com/HTML/index.html?shapiro_wilk.html) Repeated Measures ANOVA is considered to be robust against non-excessive violations of the normality requirement (Glass et al. 1972, p.273). Therefore, the results of RMA are reported even if the Saphiro-Wilks test did not designate the normality of data at all eight time levels (i.e., for all experimental sessions) or if six or seven out of eight time levels were normally distributed. However, to confirm the significance of these RMA results, **non-parametric statistics** which does not require normally distributed data was used in addition. **Friedman's Test** was used in both of the cases of non-extreme violation and extreme violation of normality. In the following table one of the results of the Saphiro-Wilks test for the parameter *Normalized Token Count of Nouns* is shown. The test indicates that the data is normally distributed for all time levels except the 7^{th} experiment.

Table 9 Saphiro-Wilks test of Normality for *Normalized Token Count of Nouns*

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Exp1	,169	8	,200*	,927	8	,491
Exp2	,164	8	,200*	,960	8	,813
Exp3	,181	8	,200*	,958	8	,787
Exp4	,133	8	,200*	,949	8	,705
Exp5	,238	8	,200*	,928	8	,494
Exp6	,195	8	,200*	,913	8	,376
Exp7	,337	8	,008	,763	8	,011
Exp8	,207	8	,200*	,873	8	,162

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

The results of the normality tests are not reported for each statistical test of parameters. If Friedman's test is used, this indicates that at least for one of the eight time levels (exp1, epx2, …, exp8) the data was not normally distributed. Except the ideal case of the normal distribution of the data for all experiments (i.e., for all eight time levels), the use of Friedman's test should be considered as a second check for confirming the significance of the change of the parameter under examination.

As it can be seen in Chapter 4, there was a 100% consistence with the results of Repeated Measures ANOVA and Friedman's Test. This can be taken as another piece of evidence of RMA's general robustness against the violation of the normality assumption. However, even moderate violations of the normality assumption for RMA in certain conditions should be kept in mind as a possible, though unlikely, limitation when the results are generalized.

3.7.4. LEXICAL CATEGORY ANALYSIS

The communication content of each couple is categorized into lexical items; lexical items are detected, counted and categorized in terms of classical

grammatical categories: Noun, Pronoun, Verb, Adj./Adv., Preposition, Interjection, Conjunction, Symbol (Radford, 1997).

3.7.4.1. CATEGORIZATION:

Most of the lexical items (words) created by each of the dyads corresponded to actual vocabulary of the shared NL or NL's (namely Turkish and English), so their categorization as noun, verb, pronoun or adjective was trivial and unproblematic. ("y", "qree" or "yeqıl" for green as adjective, "u" for you as a pronoun etc. Since they were allowed to use the letters "q, w, e, r, t, y, u, ı, o, p"; once they realized the possibility of compressing and expressing NL words in this set, they also exploited the phonetic resemblance between other un-allowed characters and the allowed ones as well as the computer-based textual communication conventions. They used q for k, capıtal ı, namely I for L, p for b, and so forth. With the given characters, a larger set of NL words than expected are created in the new communication system with minimal ambiguity and recognizable with a minimum effort (e.g., "peqıe" meaning "bekle", "wait",).

There are also communication system specific lexical items which had been used consistently by couples across the eight experiments and sometimes across all the couples. The use of the symbols → and ? was mostly deictic, or had indexical, emotional or interrogative functions that do not necessarily correspond to a unique or any NL word. In this context most of the use of symbols are categorized into interjections ("???" for "I don't know", or "what are you/we? doing now", "qup" for "the cube is there", ("küp orada işte!").

3.7.4.2. EXPLANATION FOR THE CATEGORIZATION OF THE SYMBOL "?" AS AN INTERJECTION

Lexical categories are about the functions of the words in the sentence structure, and the categorization depends on the morpho-syntactic features (Radford 1997, p. 29). Nouns name things (can be plural), verbs represent actions

(cannot be plural), lexical categories do not necessarily relate with the sentence's pragmatic function. However, having a universally acceptable and uncontroversial categorization system is considered as an impossible mission (Baker. 2003, 3). For this study, the task of categorization is mostly about the very prototypical members of categories, so is out of scope of the discussions in the literature and the main aim is to have an operational consistent criteria just limited for this study.

The use of the symbol ? is categorized as an interjection when it is used as the sign of mental states of uncertainty, curiosity or confusion. "??" was used for "I don't know", or "what are you/we? Doing now", or "I don't understand". This is unproblematic with the general definition of interjections as signs of emotions or mental states. However, as a general exception to the remaining lexical categories, the interjections are categorized regarding their pragmatic function and so "?" or "??" are categorized as interjections.

Since the symbol ? is a part of written communication of NL in general, but not a grammatical category or functional morpheme in NL, it has its own function in these emergent communication system as adding an interrogative aspect to the utterances. This use as a lexical item falls into the lexical category of interjection as a particular categorization preference for this study. The symbol ? is used as function symbol in its genuine textual use and also as a whole interrogative pronoun or function word like "what", "where" or "how", "did you?", "are you" "is it?" (e.g., "pultu??", "qre?" "is it sphere?"). This second use for this exclamatory/interrogative and pragmatic function is denoted by "? #2" in the lexical items list and categorized into interjections. The list of lexical items and their categorization for Couple 7, available at Table 10.

Table 10 List of lexical items, counts and lexical categorization of words created by Couple 7

LIST OF LEXICAL ITEMS

couple 7 (T_O)

	Lexical item	first used in exp	code	first used by	# used by p1	# used by p2	meaning	category	exp1	exp2	exp3	exp4	exp5	exp6	exp7	exp8	total # used
1	? #2	exp1	8	p2	0	1	what?, where?	intj.	1	0	0	0	0	0	0	0	1
2	y	exp1	8	p1	14	10	yes/evet	intj.	1	2	2	4	4	7	3	3	26
3	q	exp2	7	p1	9	9	küp/cube	noun	0	2	0	6	0	7	9	0	24
4	"-->"	exp2	7	p1	10	13	this is (it),is this?,there,	symbol	0	2	2	5	3	4	3	4	23
5	y #2	exp3	6	p2	13	17	yeşil/green	adj.	0	0	3	12	2	0	4	9	30
6	p	exp3	6	p2	6	9	piramit	noun	0	0	3	0	3	4	0	5	15
7	O	exp4	5	p1	9	8	küre/sphere	noun	0	0	0	6	3	0	4	4	17
8	?	exp4	5	p1	5	1	question (function)	intj.	0	0	0	3	1	2	0	0	6
9	w	exp4	5	p1	2	0	where	adv.	0	0	0	2	0	0	0	0	2
10	r	exp4	5	p1	1	0	are	verb	0	0	0	1	0	0	0	0	1
11	u	exp4	5	p1	1	0	you	pronoun	0	0	0	1	0	0	0	0	1
12	ı	exp4	5	p1	1	0	is	verb	0	0	0	1	0	0	0	0	1
13	y #3	exp4	5	p2	0	1	yakınında/yanında	adj.	0	0	0	1	0	0	0	0	1
14	t	exp5	4	p2	1	1	turuncu	adj.	0	0	0	0	0	2	0	0	2
15	tree	exp6	3	p2	0	2	ağaç/tree	noun	0	0	0	0	2	0	0	0	2
16	p #2	exp6	3	p2	5	5	purple/eflatun	adj.	0	0	0	0	0	6	4	0	10
17	e	exp6	3	p1	3	2	purple/eflatun	adj.	0	0	0	0	0	5	0	0	5
18	"-"	exp7	2				"-- --"		0	0	0	0	0	0	0	0	0
19	"-"	exp8	1				"-- --"		0	0	0	0	0	0	0	0	0
20																	
21																	
22			P1 new	10													
23			P2 new	7													
24																	

The data is summarized for the eight experiments in terms of totals and averages as it is done for the quantitative parameters in to eight different data files for the eight couples.

Table 11 A Sample Table of Token Number of Lexical Items belonging to each Lexical Category (Couple 7)

CP7	**noun**	**pro**	**verb**	**adj/v**	**prep**	**conj**	**intj**	**sym**	**Totals for the session**
exp1	0	0	0	0	0	0	2	0	2
exp2	2	0	0	0	0	0	2	2	6
exp3	3	0	0	3	0	0	2	2	10
exp4	12	1	2	15	0	0	7	5	42
exp5	6	0	0	2	0	0	5	3	16
exp6	11	0	0	13	0	0	9	4	37
exp7	13	0	0	8	0	0	3	3	27
exp8	9	0	0	9	0	0	3	4	25
Totals	56	1	2	50	0	0	33	23	

For each couple the use ratio of each lexical category is calculated for every experimental session. This is called “normalization” in the results chapter and discussed there.

Table 12 A Sample Table of the Normalized Number (Use ratios) of Lexical Items belonging to each Lexical Category (Couple 7)

CP7	noun	Pron	verb	adj/v	prep	conj	intj	Sym
exp1	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00
exp2	0,33	0,00	0,00	0,00	0,00	0,00	0,33	0,33
exp3	0,30	0,00	0,00	0,30	0,00	0,00	0,20	0,20
exp4	0,29	0,02	0,05	0,36	0,00	0,00	0,17	0,12
exp5	0,38	0,00	0,00	0,13	0,00	0,00	0,31	0,19
exp6	0,30	0,00	0,00	0,35	0,00	0,00	0,24	0,11
exp7	0,48	0,00	0,00	0,30	0,00	0,00	0,11	0,11
exp8	0,36	0,00	0,00	0,36	0,00	0,00	0,12	0,16
Totals								

In Table 12, the row for experimental session 4 (exp4) is calculated by dividing every number of nouns in the exp4 row of the Table 10, by the total number of lexical items used in that session, which is 42.

Similar explorations for the counts and use ratios of lexical categories in terms of descriptive data analysis are done as it was done for the quantitative parameters. Using the unified data tables of the eight couples (gathered from similar tables like the above Table 12 for Couple 7), the trends for the total use numbers and the use ratios of all couples are examined.

For example: The unified table of 8 Couples for the numerical trends of use ratios of nouns across the 8 experiments.

Table 13 A Sample Table of Use ratios (Normalized Number) of Nouns

Nouns Token Nrm	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8
Exp.1	0,25	0,10	0,06	0,13	0,38	0,20	0,00	0,00
Exp.2	0,24	0,00	0,08	0,07	0,17	0,16	0,33	0,10
Exp.3	0,13	0,14	0,03	0,14	0,20	0,19	0,30	0,10
Exp.4	0,17	0,20	0,11	0,14	0,19	0,28	0,29	0,25
Exp.5	0,10	0,24	0,17	0,24	0,18	0,24	0,38	0,14
Exp.6	0,10	0,10	0,14	0,09	0,27	0,20	0,30	0,22
Exp.7	0,17	0,20	0,13	0,18	0,34	0,14	0,48	0,16
Exp.8	0,15	0,23	0,13	0,15	0,37	0,18	0,36	0,27

When a graph of mean value of all eight couples is produced to see the trend, we get Figure 14:

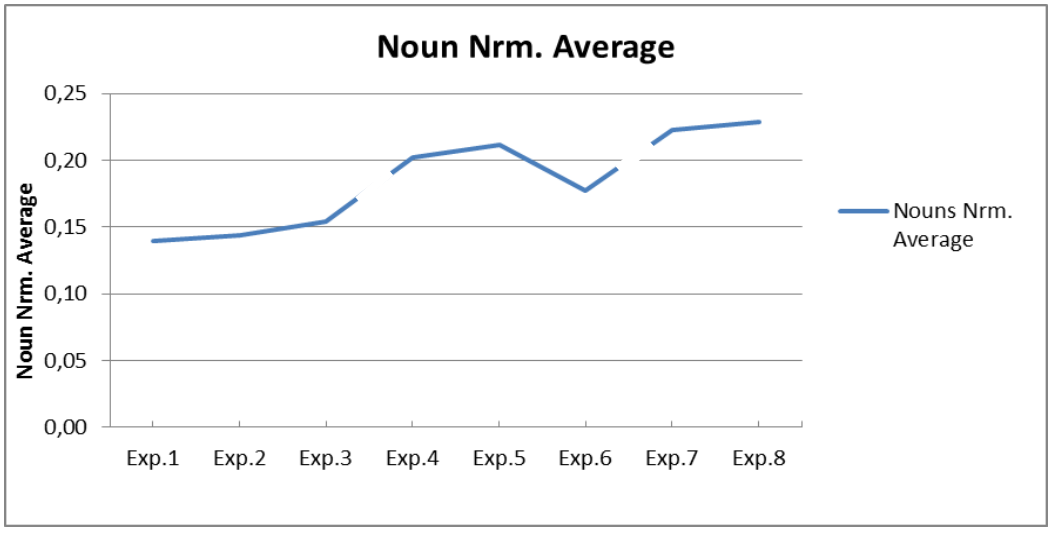


Figure 14 The Mean value of normalized values (use ratios) for nouns of each couple for the 8 experiments

The blue and green transparent columns represent the two interventions after Experiment 3 and Experiment 6.

The mean value of the normalized values for nouns of each couple for the 8 experimental sessions is plotted. The initial impression of a trend of increase in the uses ratios is scrutinized. The question of change in the use ratio of nouns is statistically tested by means of **Repeated Measures Anova**. [Give results only in the results section]For the level or lexical category analysis, similar general trends for all categories are examined and compared with each other in the results chapter.

3.7.5. SYNTACTIC COMPLEXITY

A three-level categorization is proposed for classifying the utterances of each turn in terms of their syntactic complexity, referring to similar ones used in linguistic and discursive studies of communication (Sotillo, 2000). The syntactic complexity parameter is used to observe and measure the change or difference in syntactic complexity of the utterances during the eight experimental sessions. There are three ordinal categories:

a-) **Type 1**: Single word, single symbol utterances, usually used as shortcuts or aliases for single words, simple messages and basic speech acts like “o” (for “ok”); “?” (for “I did not understand”); “y” (for “yes”); “yor” (for “mor, -violet”) etc.

b-) **Type 2:** Utterances with at least two words which contains a head and a complement (noun or verb prhrases). [expl.]

c-) **Type 3:** Utterances which contain words that can refer to a full sentence both with meaning and syntactic elements like “ı yuppıeee turu top” (for “I found the orange sphere”)

In addition, to be able to visualize the variation of sentence complexity across the eight experiments, a complexity coefficient is assigned; 1, 2 or 3 points, respectively, to quantify the total complexity in each experiment. For example, for the 5th experiment of Couple 3, if there are 12 type 1, 7 type 2 and 11 type 3 utterances, the overall complexity score is 59 (12*1+7*2+11*3=59).

The level of syntactic complexity contains as only parameter the syntactic complexity score.

The unified data file for this level of analysis is presented in Table 14.

Table 14 Syntactic Complexity Score

	Cp1	Cp2	Cp3	Cp4	Cp5	Cp6	Cp7	Cp8	Average
exp1	20,00	38,00	63,00	14,00	41,00	12,00	5,00	20,00	26,63
exp2	77,00	12,00	30,00	33,00	69,00	30,00	7,00	29,00	35,88
exp3	44,00	39,00	33,00	7,00	24,00	28,00	10,00	28,00	26,63
exp4	73,00	52,00	77,00	23,00	83,00	49,00	37,00	84,00	59,75
exp5	62,00	41,00	90,00	97,00	50,00	42,00	20,00	49,00	56,38
exp6	65,00	41,00	106,00	33,00	33,00	40,00	39,00	58,00	51,88
exp7	56,00	40,00	46,00	17,00	43,00	43,00	21,00	41,00	38,38
exp8	100,00	35,00	44,00	28,00	29,00	40,00	25,00	27,00	41,00

The presented values in Table 14 are used to examine the trend of syntactic complexity across the 8 experiments. The relation with the total amount of communication variables like total number of turns or total token number of lexical

items from the quantitative parameters level will be compared in the results chapter. If necessary, a procedure of normalization is applied to detect the trend of change factoring out the effect of communication amount.

3.7.6. SPEECH ACT ANALYSIS

The general categorization of speech acts differs from the one used in the present study, according to the point of view and scope of the study.

One categorization is in fact a list of features of speech acts by Austin (1962) and Searle (1969) which are by no means mutually exclusive. This categorization can be considered as a hierarchical one where the Utterance Act at the left-most side contains the rest (Akmajian et al. 2002, p. 395); Illocutionary, Perlocutionary, and Propositional acts are mutually exclusive, however.

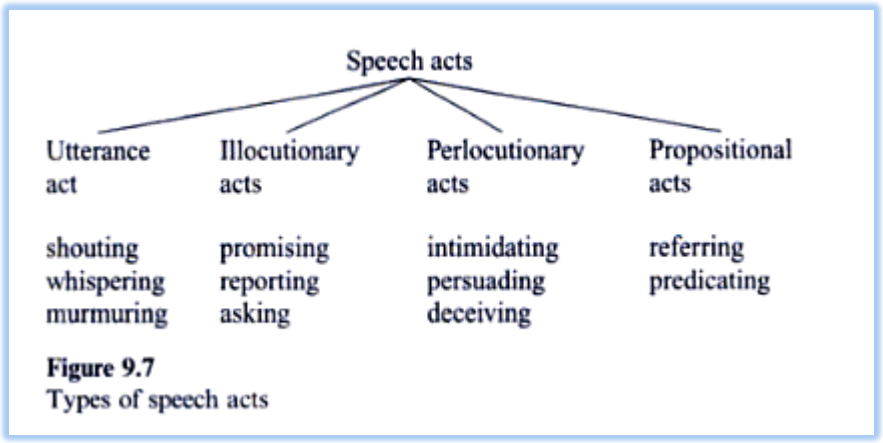


Figure 15 The Classical categories of speech acts (Akmajian et al., 2002, p. 395)

The revised categorization by Kent and Bach, which has only four speech acts, as defined in

Table 15, is adopted. (Bach, K., Routledge Encyclopedia of Philosophy, Speech Acts entry; Searle, 1969; Bach, K., 1979).

Table 15 Definitions and categories of Speech Acts

	Definition of Speech Acts
Assertive	affirming, alleging, announcing, answering, attributing, claiming, classifying, concurring, confirming, conjecturing, denying, **disagreeing**, disclosing, disputing, identifying, informing, insisting, predicting, ranking, reporting, **stating**, stipulating **(also categorized as *Constative*)**
Directive	advising, admonishing, **asking**, begging, dismissing, excusing, **forbidding**, instructing, **ordering**, permitting, **requesting**, requiring, suggesting, **urging**, warning.
Expressive	apologizing, condoling, congratulating, greeting, thanking, **accepting** (acknowledging an acknowledgment) **(also categorized as *Acknowledgment*)**
Commissive	**agreeing**, guaranteeing, inviting, **offering**, **promising**, swearing, volunteering

Each turn taking (utterance) of participants is categorized into a speech act. Experiment-wise data are collected by counting the speech acts. Similar to the procedure for the other levels of analysis, these data are summarized into a table in terms of totals and normalized totals to calculate use ratio within the experimental session by dividing each experimental sessions category totals by the total number of turns in that experiment

The categories seem to be intersecting and equivocal. Referring to the literature about speech act categories, a set of more certain instructions was prepared for categorization. The detailed guideline developed for a consistent categorization is available in Appendix C.

Assertive: (also called Constative): Stating, disagreeing, announcing and answering, affirming, usually statements about facts, which have truth value.

Directive: Utterances that cause or intend to create an action (in terms of behavior or communicative action) of the hearer, as a response: Asking someone to do (or not to do) something, ordering, asking questions (to make someone answer).

Expressive: Utterances that let the hearer know about the psychological state of the speaker or that acknowledges the hearer's mental or emotional situation,

Commissive: Utterances (commitments) about future actions which promises, offers, invites or agrees to do something.

Examples:

Yes → Commissive (accepting, agreeing on an offer or promising something) Are you coming? ("Yes, I will come") (Commissive)
Yes→Expressive (Accepting an apology, congratulation etc., also "yes, I understand" is about internal psychological condition, "ikrar") Oleey we found it!! "Yes, that's great" (Expressive)
Yes→Assertive (affirming a statement about the external world) We found the target. – "Yes, we did" (affirming, Assertive)

A second rater repeated the whole categorization procedure following the guideline. **A Kappa Inter-rater agreement test** is performed to check the used categories' appropriateness/consistency to measure or characterize speech acts. The results of this test is presented in the of Speech Acts section of the results chapter.

For each of the eight couples, data tables are created containing the number of turns categorized as Assertive, Commissive, Directive and Expressive, as in Table 16 for Couple 6:

Table 16 Speech act Categories for Couple 6 across the eight experimental sessions

Experiments	*Assertive*	*Directive*	*Expressive*	*Commissive*	TOTAL # of TURNS
Exp.1	1	6	*0*	0	7
Exp.2	5	9	1	0	15
Exp.3	3	6	3	2	14
Exp.4	12	7	1	2	22
Exp.5	7	4	3	6	20
Exp.6	9	4	3	3	19
Exp.7	6	6	4	6	22
Exp.8	7	3	6	4	20

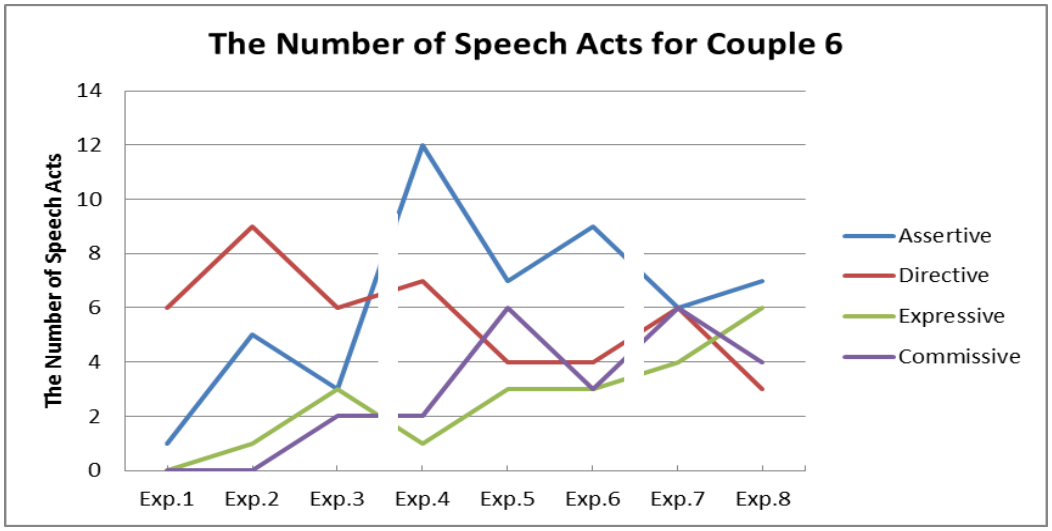


Figure 16 The Number of Speech Acts for Couple 6

The blue and green transparent columns represent the two interventions after Experiment 3 and Experiment 6

The number of turns in the four speech act categories is unified in data tables for all four categories as exemplified in the following table for Assertive Speech Acts:

Table 17 Assertive Speech Acts for the eight Couples across the eight experimental sessions

Assertive										
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	Average	Total
Exp.1	3	6	7	4	11	1	4	0	4,50	36
Exp.2	13	0	4	9	12	5	4	3	6,25	50
Exp.3	4	3	7	2	5	3	5	4	4,13	33
Exp.4	13	8	8	9	15	12	11	15	11,38	91
Exp.5	12	2	12	34	9	7	9	6	11,38	91
Exp.6	11	9	20	10	3	9	15	14	11,38	91
Exp.7	13	8	7	3	9	6	9	9	8,00	64
Exp.8	16	9	14	12	7	7	12	6	10,38	83

The above data in Table 17 allows to descriptively exploring the change of the number of Assertive Speech acts in all of the eight communication systems developed. These trends can be demonstrated in terms of averages and statistically analyzed using Repeated Measures ANOVA.

For example the trend of Assertive Speech acts in the following graph was statistically significant:

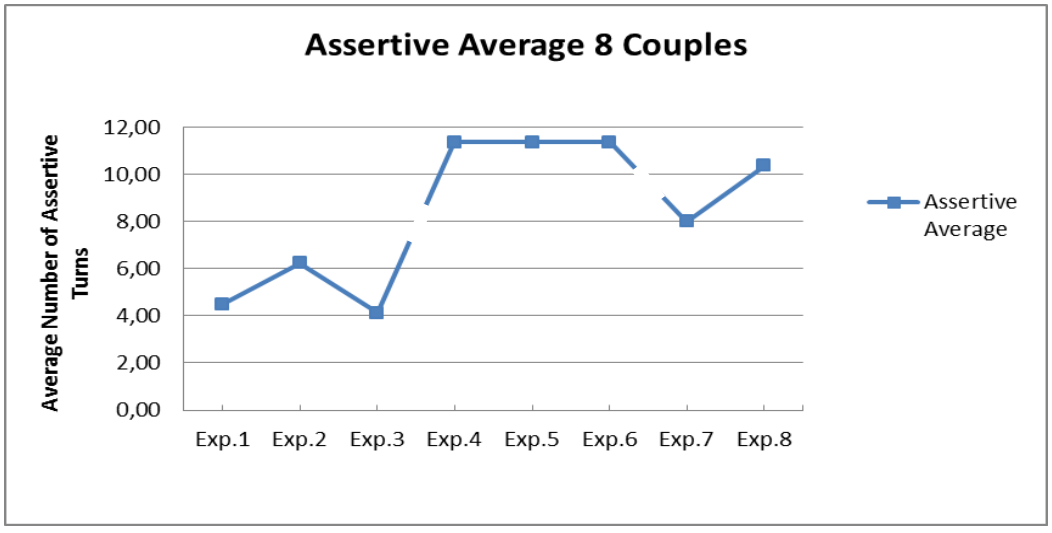


Figure 17 Average Number of Assertive Turns

The blue and green transparent columns represent the two interventions after Experiment 3 and Experiment 6

Similar exploration of the individual trends for the average numbers, total numbers and normalized numbers (use ratios) of all the remaining speech act

categories is conducted (for Commissive, Expressive and Directives, demonstrated in the results chapter). Similarly, the general trends of change for the four categories are compared in the results chapter.

3.8. GROUPS OF HYPOTHESES ON THE FOUR LEVELS OF ANALYSIS

There are three main issues concerning the interventions, manipulations or trends that have to be addressed in the construction of parameter/feature-specific hypotheses.

3.8.1. THE OVERALL CHANGE OF A PARAMETER ACROSS THE EIGHT EXPERIMENTS

There are hypotheses about the change of several variables on the 4 levels of analysis across the eight experiments. For example, quantitative parameters like **Number of Turns** in a session across the eight experiments; for the lexical categories, **use ratio of Nouns (Nouns Normalized)** or **Syntactic complexity score** or the **number of Commissive Speech acts** across the eight experiments are addressed in the hypotheses.

For example: “The turn success ratio is expected to increase across the eight experimental sessions”.

3.8.2. THE EFFECT OF INCREASING THE TARGET NUMBER IN THE FOURTH EXPERIMENT

For the first three experiments the participants received one target, however, from the 4th experiment until the last experimental session, the participants received two targets.

i-) (the change between 3rd and 4-5th experiments) (btw. 3-4,5)

The main assumption for this type of hypotheses is that in the 3-4-5 series of experiments, we expect to see an increase in the total amount of communication in terms of quantitative parameters like an increase in turn numbers, total token

number of words of the 4th and 5th sessions. Similarly, we expect a change in the lexical or speech act parameters that are affected from the amount of communication, like the number of speech acts in any category or number of nouns.

For example: "Number of Turns in a session is expected to increase in the 4th and 5th with respect to the 3rd experiment.

ii-) The development of the communication system in the uniform phase (in terms of task numbers and their difficulty) in the last 5 experiments (4-5-6-7-8th experiments).

For example "The total number of turns across the last 5 experiments is expected to decrease."

The main assumption for this type of hypotheses is that in the 4-5-6-7-8 series of experiments, the qualitative difficulty and numbers of tasks are constant. So the effect of continuous exposition to the environment and similar task are expected to produce and effect a change in the parameters that may speak for the efficient use of the communication system or the efficiency of the communication system.

For example: The total number of Assertive Speech Acts will increase in the 4-5-6-7-8 series of experiments.

3.8.3. THE EFFECT OF THE TWO MONTHS GAP BETWEEN THE 6TH AND LAST TWO EXPERIMENTS (BTW. 6-7,8TH)

The two months gap between the 6th and 7th experiments divides the development of the communication system between the 1-2-3-4-5-6 and the 7-8 series of experiments.

The main assumption or the question behind this break was to see whether the lexical content and other features of the communications systems are affected from the memory decay and whether the retrieval of the lexical contents,

navigational and communicative skills developed during the 1-2-3-4-5-6th sessions is affected.

The assumption may lead to expectations of longer series of communication for coordinating the information sharing or navigation. Similarly, if a decay in the number of lexical items is assumed, participants are expected to re-create lots of new lexical items in the 7th experiment that they had already created and converged on as established conventions.

Examples:

The total amount of turns will increase after the 6th, i.e., in the 7th and 8th experimental sessions.

The number of new lexical items created in the 7th and 8th experimental session will exceed that in the 6th one.

3.9. PRODUCTION OF ACTUAL HYPOTHESES

The trend of the change for any parameter can be hypothesized and examined in accordance with the above segmentation of the three experimental series, (1 to 8th), (3-4,5/4,5,6,7,8) and (6-7,8). For instance, the change for the quantitative parameter **Total number of turns** can be hypothesized as follows:

i-) (1 to 8th)

The total number of turns in each experimental session across the eight experiments is expected to increase (1st to 8th experiments)

ii-) (3-4,5)

The total number of turns in each experimental session after the transition from the single target to the two targets version is expected to increase (increase between the 3rd and 4th experiment).

iii-) (4,5,6,7,8)

The total number of turns in each experimental session during the two targets phase is expected to decrease (decrease between the 4th and 8th experiment)

iv-) (6-7,8)

The total number of turns in each experimental session after the two months break is expected to increase (increase between the 6th and 7th experiment)

For each of the 24 quantitative parameters (and some additional combinations of them in terms of ratios), the 9 lexical categories, 4 speech act parameters and syntactic complexity measures 4 sets of hypotheses can be developed. However, in this section, such an exhaustive production of hypotheses will not be carried out. Rather, in the following section some general hypotheses that have been developed during the initial data analysis of the pilots will be listed. However, in the later phases of the data analysis, several other significant observations and inferences are made and tested.

3.10. LIST OF HYPHOTHESES

A- Quantitative Parameters

Number of Turns in each experimental session

Overall change across the eight experiments

1- H1: Number of Turns in each experimental session will change across the eight experimental sessions.

Change after the 4th session intervention

2- H1: Number of Turns will change across 3rd, 4th and 5th experimental sessions.

Change across the 4-5-6-7-8th sessions period.

3- H1: Number of Turns will change across the 4-5-6-7-8th experimental sessions.

Change after the two months gap (between the 6th and last two experiments)

4- H1: Number of Turns will change across the 6th, 7th and 8th experimental sessions.

This set of hypotheses predicts an increase in Number of Turns in the initial sessions and a relative decrease at the later sessions, as are compared to Number of Turns at intermediate sessions (like the 4th or the 5th experiments)

Turn Success Ratio:

Overall change across the eight experiments

5- H1: Turn Success Ratio will change across the eight experimental sessions.

Change after the 4th session intervention

6- H1: Turn Success Ratio will change across the 3rd, 4th and 5th experimental sessions.

Change across the 4-5-6-7-8th sessions period

7- H1: Turn Success Ratio will change across the 4-5-6-7-8th experimental sessions.

Change after the two months gap (between the 6th and last two experiments)

8- H1: Turn Success Ratio will change across the 6th, 7th and 8th experimental sessions.

This set of hypotheses predicts an increase in Turn Success Ratio, where the ratio converges to 1 after the initial sessions.

Token Number of Lexical Items

Overall change across the eight experiments

9- H1: Token Number of Lexical Items will change across the eight experimental sessions.

Change after the 4th session intervention

10- H1: Token Number of Lexical Items will change across the 3rd, 4th and 5th experimental sessions.

Change across the 4-5-6-7-8th sessions period

11- H1: Token Number of Lexical Items will change across the 4-5-6-7-8th experimental sessions.

Change after the two months gap (between the 6th and last two experiments)

12- H1: Token Number of Lexical Items will change across the 6th, 7th and 8th experimental sessions.

This set of hypotheses predicts an increase in Token Number of Lexical Items in the initial sessions and a relative decrease at the later sessions, when they are compared to Token Number of Lexical Items of intermediate sessions (like 4th or the 5th experiments)

Number of New Lexical Items

Overall change across the eight experiments

14- H1: Number of New Lexical Items will change across the eight experimental sessions. What about the other two hypotheses? If you do not hypothesize them, maybe you tell below, why not

Change after the two months gap (between the 6th and last two experiments)

15- H1: Number of New Lexical Items will change across the 6th, 7th and 8th experimental sessions.

This set of hypotheses predicts a continuous decrease in the number of new lexical items after the initial sessions where values will be high.

Type Number of Lexical Items

Overall change across the eight experiments

16- H1: Type Number of Lexical Items will change across the eight experimental sessions.

This hypothesis estimates an increase in Type Number of Lexical Items in the initial sessions and a relative decrease at the later sessions, when compared to Type Number of Lexical Items at intermediate sessions (like the 4th or the 5th experiments)

TYPE over TOKEN Number of Lexical Items ratio:

Overall change across the eight experiments

17-H1: TYPE over TOKEN Number of Lexical Items ratio will change across the eight experimental sessions.

Change after the 4th session intervention

18- H1: TYPE over TOKEN Number of Lexical Items ratio will change across the 3rd, 4th and 5th experimental sessions.

Change across the 4-5-6-7-8th sessions period

19- H1: TYPE over TOKEN Number of Lexical Items ratio will change across the 3rd, 4th and 5th experimental sessions.

Change after the two months gap (between the 6th and last two experiments)

20- H1: TYPE over TOKEN Number of Lexical Items ratio will change across the 6th, 7th and 8th experimental sessions.

This set of hypotheses predicts an increase in TYPE over TOKEN Number of Lexical Items ratio after the initial sessions, when some adaptation is achieved. A decrease is expected when the adaptation is not achieved or there is a need for re-adaptation.

Turn Taking ratio p1/p2:

Overall change across the eight experiments

21- H1: Turn Taking ratio p1/p2 will change across the eight experimental sessions.

This set of hypotheses estimates an increase in Turn Taking ratio p1/p2, where the ratio converges to 1 after the initial sessions.

B- Syntactic Complexity Score and Normalized Syntactic Complexity Score:

Overall change across the eight experiments

22- H1: The Syntactic Complexity score in each experimental session will change across the eight experimental sessions.

23- H1: The normalized Syntactic Complexity score in each experimental session will change across the eight experimental sessions.

This set of hypotheses predicts an increase in the Syntactic Complexity score across the eight experiments. The verification of the change in the normalized Syntactic Complexity score aims to avoid the interference due to the amount of communication.

C- Lexical Category

Token and Normalized Number of Nouns

Overall change across the eight experiments

24- H1: Token Number of Nouns in each experimental session will change across the eight experimental sessions.

25- H1: Normalized Number of Nouns in each experimental session will change across the eight experimental sessions.

This set of hypotheses predicts an increase in Token and Normalized Number of Nouns across the eight experiments.

Token and Normalized Number of Pronouns

Overall change across the eight experiments

26- H1: Token Number of Pronouns in each experimental session will change across the eight experimental sessions.

27- H1: Normalized Number of Pronouns in each experimental session will change across the eight experimental sessions.

This set of hypotheses predicts a decrease in Token and Normalized Number of Pronouns across the eight experiments.

Token and Normalized Number of Verbs

Overall change across the eight experiments

28- H1: Token Number of Verbs in each experimental session will change across the eight experimental sessions.

29- H1: Normalized Number of Verbs in each experimental session will change across the eight experimental sessions.

This set of hypotheses predicts an increase in Token and Normalized Number of Verbs across the eight experiments.

Token and Normalized Number of Adjectives and Adverbs

Overall change across the eight experiments

30- H1: Token number of Adjectives and Adverbs in each experimental session will change across the eight experimental sessions.

31- H1: The normalized number of adjectives and adverbs in each experimental session will change across the eight experimental sessions.

This set of hypotheses predicts an increase in Token and Normalized Number of Adjectives and Adverbs across the eight experiments.

Token and Normalized Number of Interjections

Overall change across the eight experiments

32- H1: token Number of Interjections in each experimental session will change across the eight experimental sessions.

33- H1: Normalized Number of Interjections in each experimental session will change across the eight experimental sessions.

This set of hypotheses predicts an increase in Token and Normalized Number of Interjections across the eight experiments.

D- Speech Act Category Analysis

Number and Normalized Number (use ratio) of Assertive Speech Acts

Overall change across the eight experiments

34- H1: Number of Assertive Speech Acts in each experimental session will change across the eight experimental sessions.

35- H1: Normalized Number (use ratio) of Assertive Speech Acts in each experimental session will change across the eight experimental sessions.

This set of hypotheses predicts an increase in Number and Normalized Number of Assertive Speech Acts across the eight experiments.

Number and Normalized Number (use ratio) of Directive Speech Acts

Overall change across the eight experiments

36- H1: Number of Directive Speech Acts in each experimental session will change across the eight experimental sessions.

37- H1: Normalized Number (use ratio) of Directive Speech Acts in each experimental session will change across the eight experimental sessions.

This set of hypotheses predicts a decrease in Number and Normalized Number of Directive Speech Acts across the eight experiments.

Number and Normalized Number (use ratio) of Expressive Speech Acts

Overall change across the eight experiments

38- H1: Number of Expressive **S**peech Acts in each experimental session will change across the eight experimental sessions.

39- H1: Normalized Number (use ratio) of Expressive Speech Acts in each experimental session will change across the eight experimental sessions.

This set of hypotheses does not predict a direction in change.

Number and Normalized Number (use ratio) of Commissive Speech Acts

Overall change across the eight experiments

40- H1: Number of Commissive Speech Acts in each experimental session will change across the eight experimental sessions.

41- H1: Normalized Number (use ratio) of Commissive Speech Acts in each experimental session will change across the eight experimental sessions.

This set of hypotheses does not predict a direction in change.

CHAPTER 4

4. THE DESCRIPTIVE AND STATISTICAL ANALYSIS OF THE EXPERIMENTAL DATA AND QUALITATIVE RESULTS

As presented in the method chapter the communicative data is analyzed in the following groups of parameters (or levels of analysis):

1- Quantitative and temporal Analysis
2- Syntactic Complexity Analysis
3- Lexical Category Analysis
4- Speech Act Analysis

The detected trends of change for the parameters will be expressed as increases or decreases in the interpretations of the data. It is important to keep in mind that all these increases, decreases and differences are between the scores of different experimental sessions, not within sessions. Consequently the interpretations of the graphs representing the change across experimental sessions should be based on this fact.

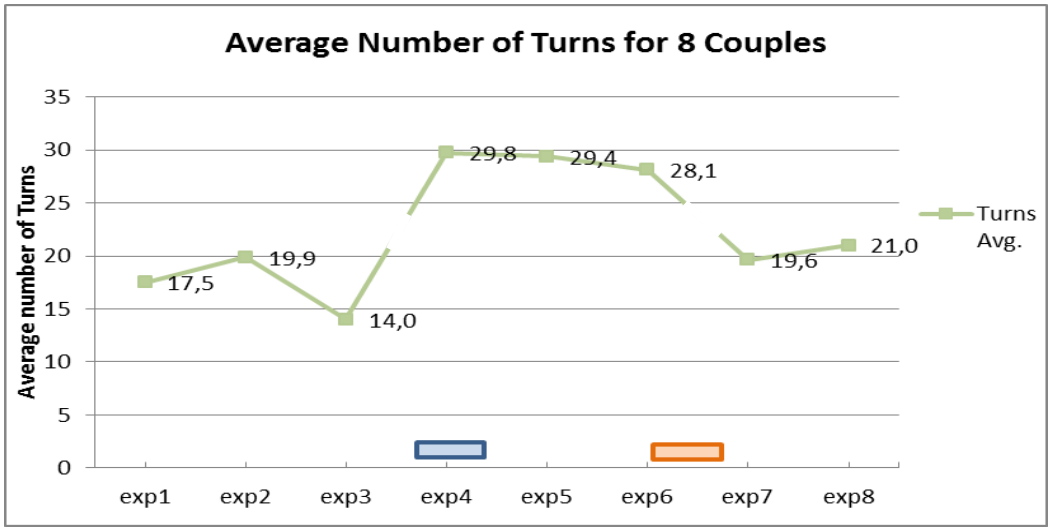


Figure 18 Average Numbers of Turns for 8 Couples

A note on reading the following Figures: In Figure 18, although the format of a line-graph has been chosen to indicate the time passing from exp1 to exp8, nothing happens between the time points *exp1* and *exp2*, that is, the line segment connecting the values 17.5 and 19.9 is just for visualizing the difference (the increase) between the averages of exp1 values and exp2 values. The entire green series line should be interpreted accordingly.

Please further note that at two time points, between exp3 and exp4 and between exp6 and exp7, two vertical colored bars, one blue, one orange, have been added, respectively. These vertical bars indicate some general change in the experimental series: (1) the blue bar indicates a change in the instruction (finding one target *vs.* finding 2 targets), (2) the orange bar indicates a temporal gap of about 2 months between those two sessions. At these points in time, two crucial manipulations in the experimental setting were made, allowing the observation of changes in parameters contingent on these manipulations. To further emphasize these two manipulations in all of the line charts, the trend lines connecting the value points of exp3 - exp4 and exp6 - exp7 sketched with a hole. For example,

taking only the first hole between exp3 and exp4 into consideration, the remaining five experiments (4-5-6-7-8th) can be visualized in mind as a separate segment of experimental sessions with two targets. Same holds to visualize to two months gap between 6th and 7-8th experiments.

After the analysis of the quantitative parameters, the term “normalization” is introduced for the lexical category, syntactic complexity and speech act category analysis. It is not used in its strict statistical terminological sense. Normalization is a standardizing procedure its formulas serve the purpose of removing scaling effects or errors. In the context of this study, it is used to denote the operation of removing the effects of one variable to the other. For example, variation in the amount of communication (number of turns, total number of lexical items) for a series of experimental sessions may make it harder to visualize the variance or distribution counts of categories of some other variable. The change in the number of nouns or the change in the counts of turns categorized into assertive speech acts may be affected by the change in the general amount of communication. In such cases, for example, to compare the increase in the numbers of nouns with the change in the numbers of lexical items in other lexical categories (verbs, pronouns), a process of normalization is applied, by dividing the score for a category by the total of the scores of all categories. The normalized number of Assertive speech acts is thus calculated by dividing the number of Assertive speech acts by the total number of turns in that session (the total number of turns is equal to total number of all categories of speech acts counts). This operation for counts of similar categorical parameters can be alternatively defined as the calculation of “use ratios” for that category. “Normalized token count of nouns” means the “use ratio of nouns” with respect to the all lexical categories.

In the last sections of the results chapter, qualitative observations about the emerging communication system are presented. These include strategies of simulating NL language under experimental limitations, couple-wise descriptions of

strategies and performance, discussion of the effects of the interventions for the 3-4-5th sessions, and 6-7th sessions, main action schemes used in the experimental sessions, the segmentation of the sessions by means of these action-scripts, scenarios, etc.

4.1. QUANTITATIVE AND TEMPORAL ANALYSIS

4.1.1. TIME

The time parameter is not analyzed because no strict maximum time limit had been imposed on dyads. It was not emphasized as a performance criterion in the instructions. However, Figure 19 and
Figure 20 may suggest a trend of task completion times of the dyads.

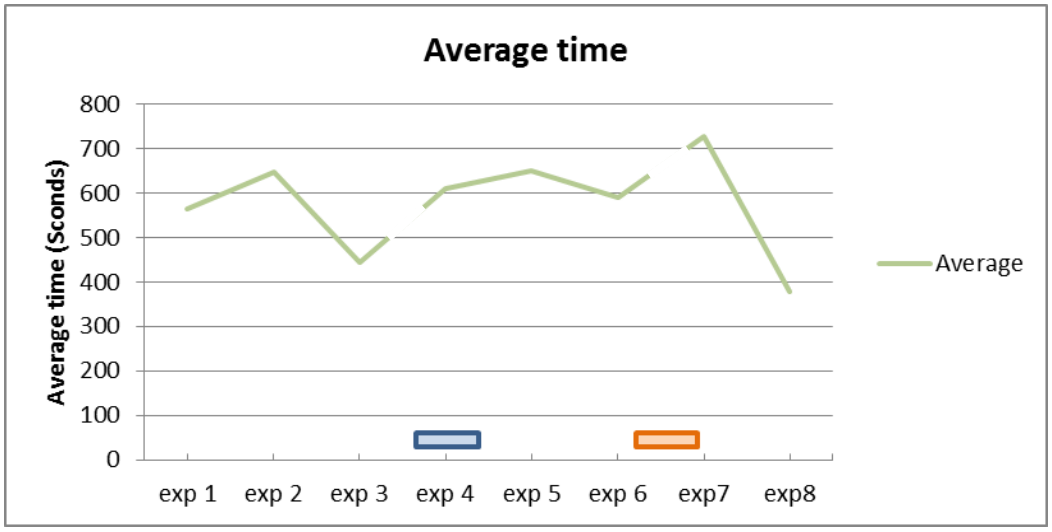


Figure 19 Average Experimental Session Time for 8 Couples across 8 experiments

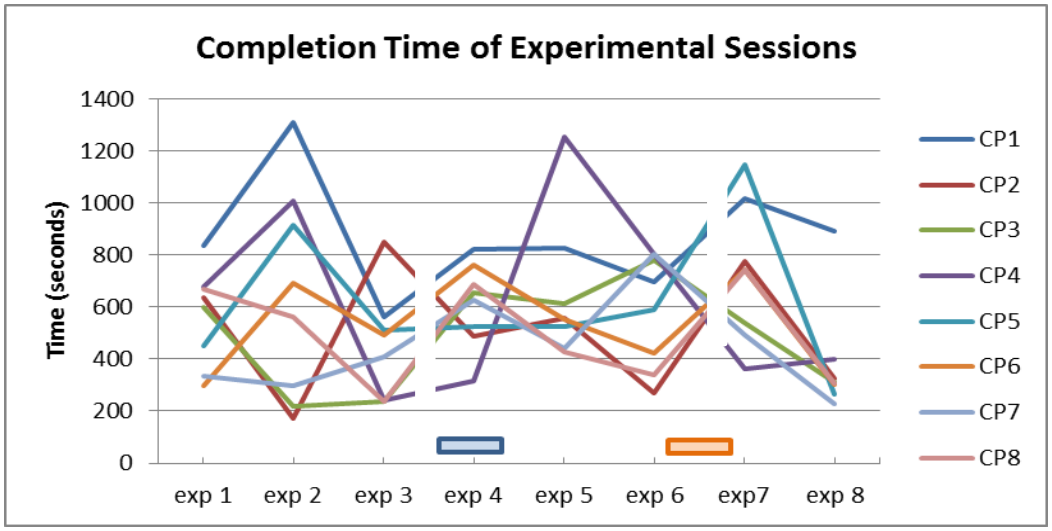


Figure 20 Experimental Session Time for 8 Couples across 8 experiments

4.1.2. TOTAL NUMBER OF TURNS:

"The total number of turns" is the total count of turn takings of the members of the dyad in an experimental session. The parameter varies according to the level of development of the communication system, the efficiency of navigation and task accomplishment performance.

The initial observations in the pilot studies proposed three hypotheses stating an overall increase is expected across the eight experiment but a slight decrease in the later sessions due to the increase in the efficiency of the communicative skills and communication system, which provides a more precise communication and coordination (that means less need of lengthy sessions of turn takings).

In the following reports of Repeated Measures ANOVA testing the development of a parameter across the 8 experiments, different numbers of experimental sessions will be analyzed in order to address the two basic manipulations discussed before:

(1) All 8 experiments will be analyzed, in order to know whether a general change occurred.

(2) Exp3, 4, and 5 will be analyzed in order to determine the effect of the first manipulation at exp4 (change in instruction).

(3) All remaining experiments after the 4th one will be analyzed, in order to know whether any change in the remaining time series occurred after the first manipulation at exp4

(4) Exp6, 7, and 8 will be analyzed in order to determine the effect of the second manipulation between exp6 and exp7 (temporal gap).

4.1.2.1. OVERALL CHANGE IN THE NUMBER OF TURNS

The change of the total number of turns across the eight experimental sessions was examined with a Repeated Measure ANOVA. It was found **significant** under the Greenhouse-Geisser correction ($F(2.9, 20.9)=3.29$, $p=.043$, $\eta_p^2=0.320$).

This significant result is due to a significant quadratic trend ($F(1, 7)=12.02$, $p=.01$, $\eta_p^2=0.62$) and a significant quartic (fourth order) polynomial trend ($F(1, 7)=5.92$, $p=.045$, $\eta_p^2=0.46$).

According to Friedman's Test, there was a statistically significant change in the total number of turns across the eight experiments ($\chi^2(7) = 25.449$, $p = .001$).

4.1.2.2. CHANGE IN THE NUMBER OF TURNS AFTER THE 4TH SESSION INTERVENTION

The change of the total number of turns across the 3rd, 4th and 5th experimental **sessions was not significant** under the Greenhouse-Geisser correction ($F(1.1, 7.9)=4.55$, $p=.063$, $\eta_p^2=0.392$).

4.1.2.3. THE CHANGE IN THE IN THE NUMBER OF TURNS IN THE 4-5-6-7-8TH SESSIONS PERIOD

The change of the number of turns across the 4-5-6-7-8th experiments **was not significant** ($F(4, 28)=1.940$, $p=.131$, $\eta_p^2=0.217$)).

4.1.2.4. THE CHANGE IN THE IN THE NUMBER OF TURNS AFTER THE TWO MONTHS GAP (BETWEEN THE 6TH AND LAST TWO EXPERIMENTS)

The change of the number of turns across the 6th, 7th and 8th experiments **was not significant** ($F(2, 14)=3.23$, $p=.070$, $\eta_p^2=0.316$)).

4.1.2.5. DISCUSSION

Participants are not able to talk much in the first session, since they do not have any shared lexical items and conventions yet. Then when they figure out to communicate in the first or second sessions, there is an increase in the communication activity reflected in the turn numbers. This expectation is verified by the test in 4.1.2.1.

However, when the strategy is settled and the dyad adapts to the constraints of the communication system, the efficient use of the emerging communication system requires fewer turns to be taken in the conversation, and a respective decrease for the parameter in the third experiment is observed.

Then the instructions are changed and the phase of having to search 2 targets begins, which obviously requires additional information sharing and collaboration to find the second target. This increase in the number of the tasks is hypothesized to merely have an effect on the relevant quantitative parameters which describe *the amount of communication* – total number of turns being one of them. So *an increase in the number of turns in the 4th experiment is expected*. The numerical increase in the descriptive data is marginally significant in Repeated Measures ANOVA in 4.1.2.2.

Similar to the decrease observed in the adaptation of the dyad in the 3rd session, there is again a decrease in the subsequent sessions due to achieving the re-adaptation to the new situation. So a decrease in the later phases of 4-5-6-7-8th sessions is expected where there are no new challenges that require any further re-adaptation but the increase in the efficiency leads to more precise communication

that requires a lower number of repetitions, read-backs or redundant utterances. The Repeated Measures ANOVA test in 4.1.2.3 for observed decrease for this period was not significant, however.

The two months gap between the first six and the 7th and 8th experimental sessions challenges the resilience of the developed communication system in the long term memories of the dyad. The expectation of an increase due the decayed efficiency should be tested, therefore. However, the Repeated Measures ANOVA test in 4.1.2.5 for the observed decrease for this period was not significant.

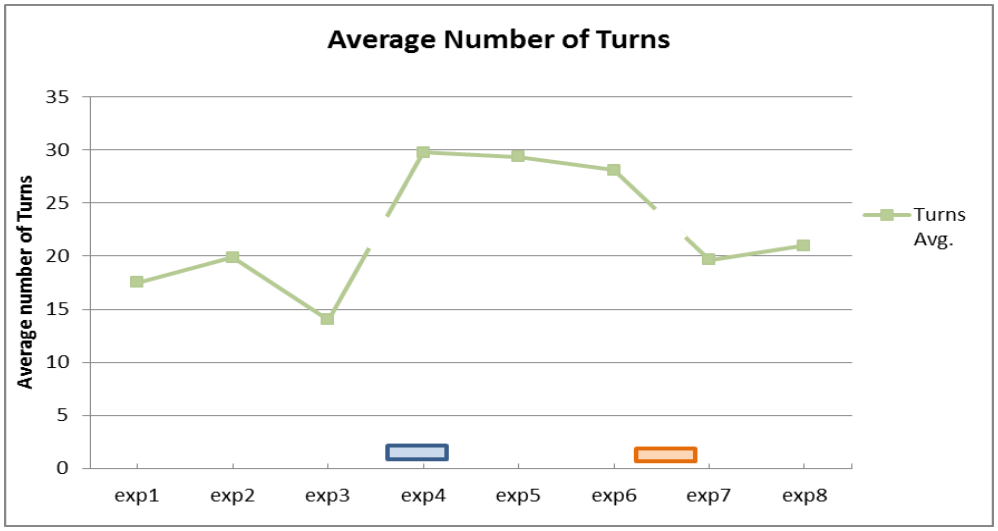


Figure 21 Average Number of Turns of all Couples across 8 experiments

4.1.3. TURN SUCCESS RATIO

The turn success ratio is the proportion of successful turns over all turns in an experimental session and shows a strong trend of increase towards the ratio 1 within the first 3 experiments, which means a complete match between the speaker meanings and the comprehension of the utterances during an experimental session.

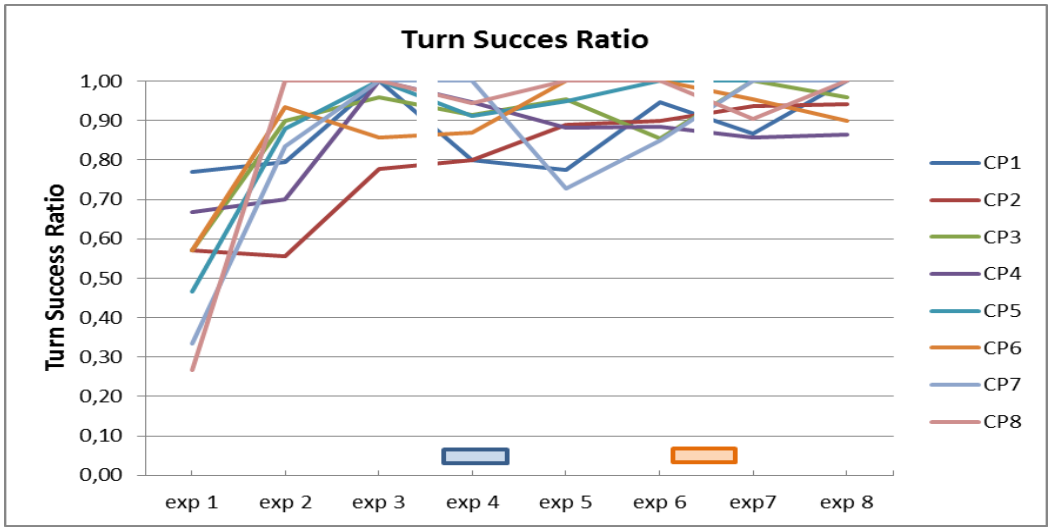


Figure 22 Turn Success Ratio for the 8 couples across the 8 experiments

4.1.3.1. OVERALL CHANGE IN THE TURN SUCCESS RATIO

The change of turn success ratio across the eight experimental sessions was **significant** ($F(7, 49)=16.12$, $p<.0001$, $\eta_p^2=0.967$).

This significant result is due to a linear trend ($F(1, 7)=55.55$, $p<.0001$, $\eta_p^2=0.88$), a quadratic trend ($F(1, 7)=20.95$, $p=.003$, $\eta_p^2=0.75$), a cubic trend ($F(1, 7)=11.66$, $p=.011$, $\eta_p^2=0.63$) and a significant quartic (fourth order) polynomial trend ($F(1, 7)=6.19$, $p=.042$, $\eta_p^2=0.47$).

According to Friedman's Test, there was a statistically significant change in the turn success ratio across the eight experiments ($\chi^2(7) = 27.037$, $p < .0001$).

4.1.3.2. CHANGE IN THE 4-5-6-7-8TH AND 6-7-8TH PERIODS

The Repeated Measure ANOVA tests on the parameter on the following critical segments of the 8 experiments period revealed that the changes were not significant:

i-) The change in the turn success ratio before and after the 4th session intervention **was not significant**

ii-) The change in the turn success ratio in the 4-5-6-7-8th sessions period **was not significant**.

iii-) The change in the turn success ratio after the two months gap **was not significant**.

4.1.3.3. DISCUSSION

The expectation of convergence of turn success ratio towards 1 is achieved generally around the 2nd and 3rd experiments. The corresponding hypothesis is verified.

The turn success ratio was not influenced by the exp4 intervention or the two months break, see Figure 23.

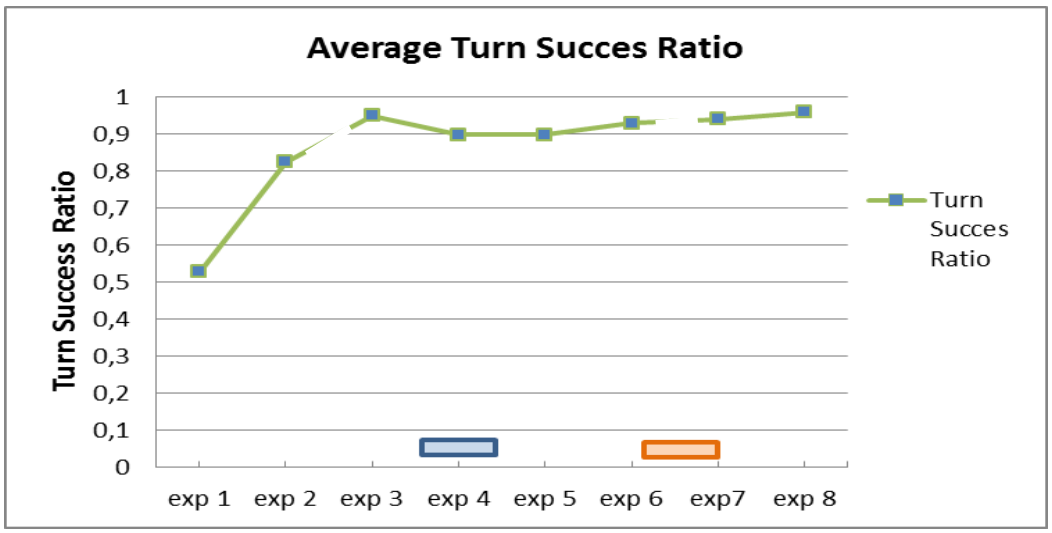


Figure 23 Average Turn Success Ratio of the 8 couples across the 8 experiments

4.1.4. TOKEN NUMBER OF LEXICAL ITEMS:

This parameter shows a great variation among couples and during the eight experiments period. A relative decrease after the first sessions and a universal increase when the experiment shifted to the “two target mode”, and after this, the total (token) number of lexical items used in each experimental session tends to

decrease due to the efficiency of emerged communication system and the communicative skills of the participants in using it. These observations are in parallel with the case number of turns in (4.1.2).

4.1.4.1. OVERALL CHANGE IN THE TOKEN NUMBER OF LEXICAL ITEMS

The change of ***token number of lexical items*** across the eight experimental sessions was **significant** ($F(7, 49)=5.40$, $p<.0001$, $\eta_p^2=0.435$). This significant result is due to a linear trend ($F(1, 7)=8.68$, $p=.022$, $\eta_p^2=0.55$), a quadratic trend ($F(1, 7)=26.53$, $p=.001$, $\eta_p^2=0.79$), and a cubic trend ($F(1, 7)=11.66$.

According to Friedman's Test, there was a statistically significant change in the token number of lexical items across the eight experiments ($\chi^2(7) = 32.733$, $p < .0001$).

4.1.4.2. CHANGE IN 3-4-5TH EXPERIMENTS PERIOD

The change in the token number of lexical items after the 4th session intervention was **significant** ($F(2, 14)=7.48$, $p=.006$, $\eta_p^2=0.516$). This significant result is due to a linear trend ($F(1, 7)=9.8$, $p=.017$, $\eta_p^2=0.583$).

4.1.4.3. CHANGE IN 4-5-6-7-8TH EXPERIMENTS PERIOD

The change in the token number of lexical items in the 4-5-6-7-8th sessions period was **marginally significant**.

The trend after the *second target mode* is introduced, for 4-5-6-7-8 is a steady decreasing one, which is verified by the Repeated Measures ANOVA with **marginal significance** ($F(4, 28)=2.60$, $p=.058$, $\eta_p^2=0.271$). This result is due to a significant linear trend ($F(1, 7)=8.53$, $p=.022$, $\eta_p^2=0.55$)

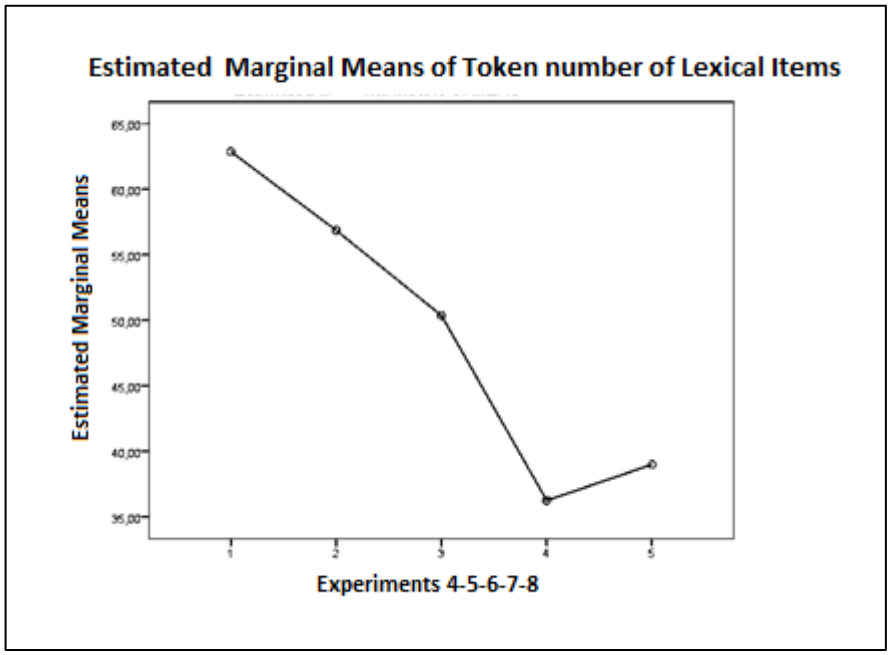


Figure 24 Estimated Marginal Means of Token number of Lexical Items for all Couples across Exp4-5-6-7-8

This secondary trend of decreasing token number of lexical items will be discussed as an indication of increase in the skill of using the emergent communication system which reduces redundant turn taking, utterances and enables some more concise communication.

There is a more clear-cut visualization of this decrease when **Couple 1 is considered as an outlier** on reasonable grounds. Couple 1 is the only dyad that used more numbers of lexical items in the 8th experiment in comparison with the 7th. This was due to an idiosyncratic problem of coordination and navigation in the environment that took a lot of time to complete the session (mostly because of the fatigue due to the length of the experiment).

When the Couple 1 data is removed, the Repeated Measures ANOVA test shows that the decrease in the 4-5-6-7-8th sessions period was significant ($F(7, 42)=6.68$, $p<.001$, $\eta_p^2=0.527$).

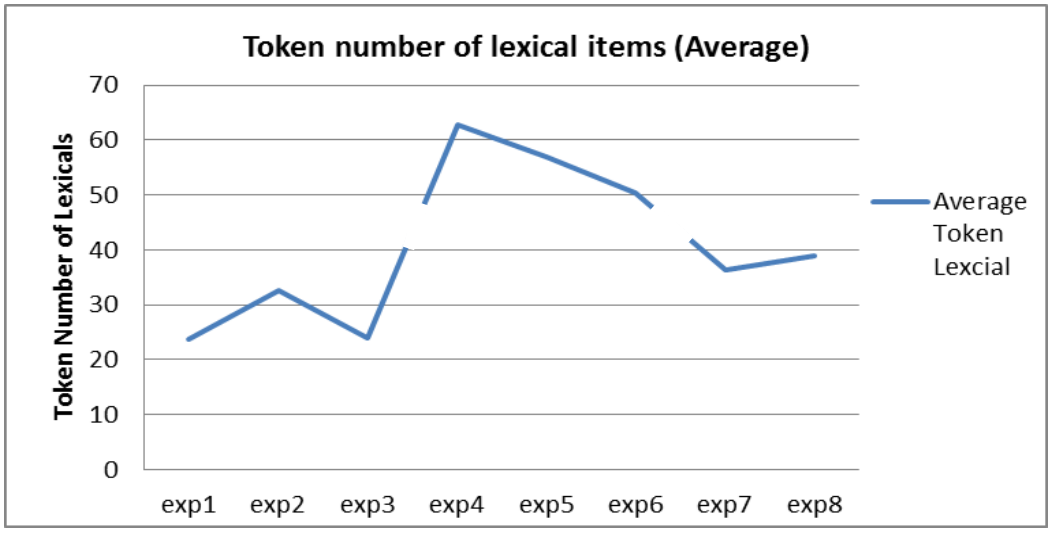


Figure 25 Average Token Number of lexical items (across all 8 couples)

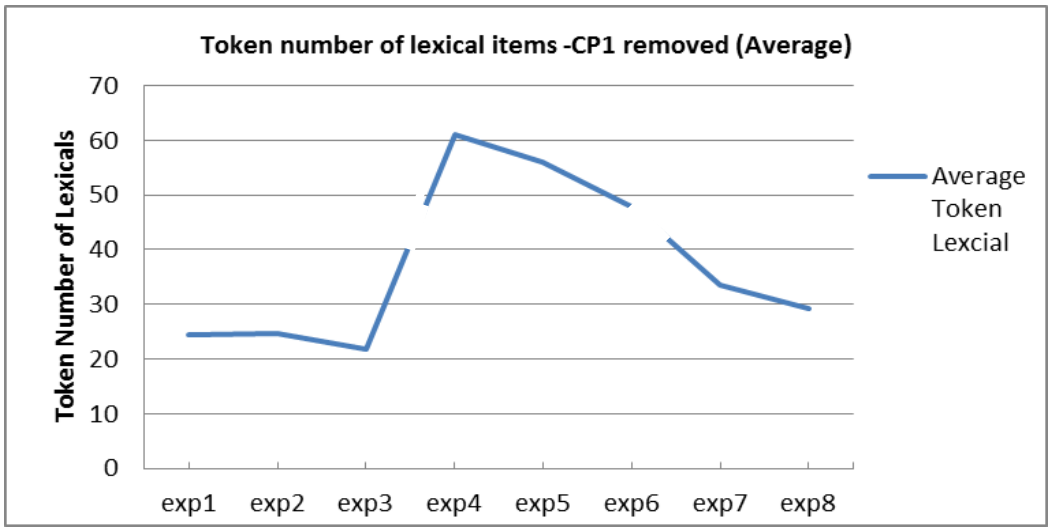


Figure 26 Average Token number of lexical items (across all couples except CP1)

4.1.4.4. CHANGE IN THE 6-7-8TH PERIODS

The change in the token number of lexical items after the two months gap was not significant. ((F (2, 14)=1.79, p=.202, η_p^2=0.204))

4.1.4.5. DISCUSSION

The expected overall increase in the average token numbers of lexical categories across the eight experiments is verified. The slight decrease in the later sessions due to the increase in the efficiency of the communicative skills and communication system is verified as well. Improved efficiency enables more precise and concise communication (that means less need of numerous lexical items).

4.1.5. NUMBER OF NEW LEXICAL ITEMS

This parameter is the number of new lexical items which are successfully introduced into the communication system in each experimental session.

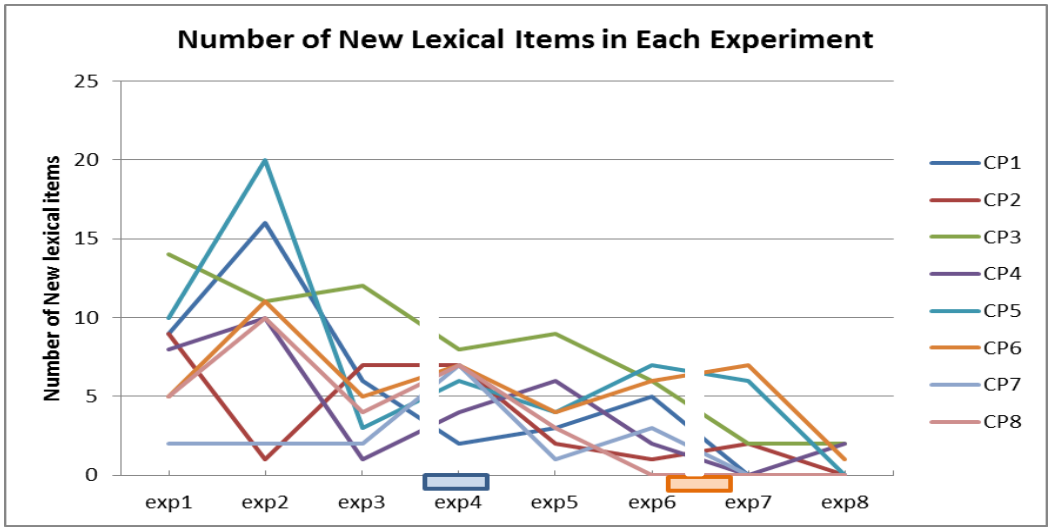


Figure 27 Number of new lexical items for the 8 couples across the 8 experiments (The blue and orange transparent bands represent the two interventions after Experiment 3 and Experiment 6)

4.1.5.1. OVERALL CHANGE IN THE NUMBER OF NEW LEXICAL ITEMS

The decrease is depicted in the descriptive analysis and tested statistically. The change of ***number of new lexical items*** across the eight experimental sessions

was **significant** (F (7, 49)=8.36, p<.0001, η_p^2=0.544). This significant result is due to a linear trend (F (1, 7)=38.07, p<.0001, η_p^2=0.85).

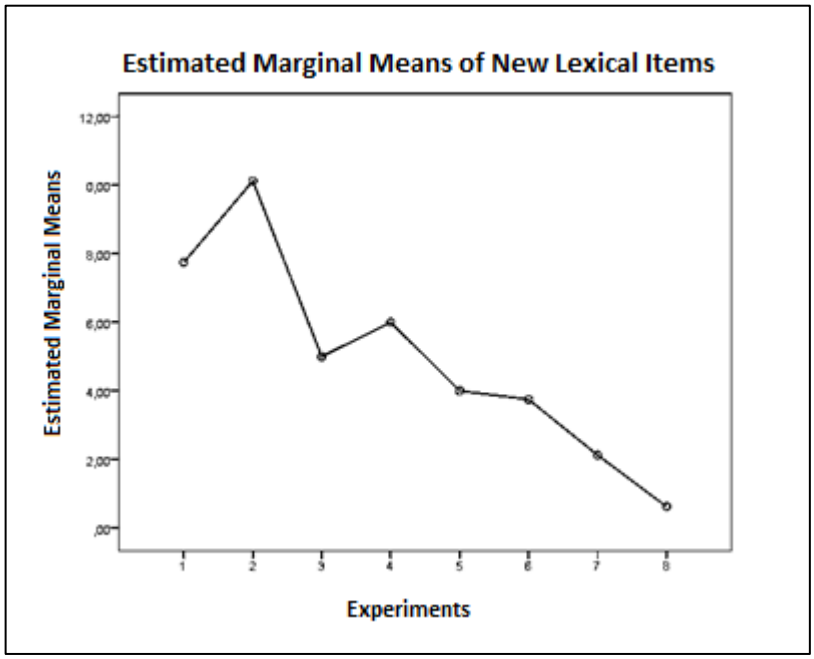


Figure 28 Estimated marginal means plot for the number of new lexical items in the RMA

According to Friedman's Test, there was a statistically significant change in the number of new lexical items across the eight experiments ($\chi^2(7) = 28.728$, $p < .0001$).

4.1.5.2. THE CHANGE IN THE NUMBER OF NEW LEXICAL ITEMS AFTER THE TWO MONTHS GAP (FOR THE 6TH, 7TH AND 8TH EXPERIMENTS PERIOD)

This continuation of decrease in the 7th and 8th experiments can be used in the later discussion for ruling out the effect of time (the 2 months gap between first 6 and last two experiments) for the persistence of lexical items in the system. Similarly it can be used to argue in favor of the persistence of the environment and task-specific communicative skills of the participants as stabilizing factors of the communicative system.

The change in the number of new lexical items across the 6th, 7th and 8th experiments was **significant** (F (2, 14)=5.67, p=.016, η_p^2=0.448)). This significant result is due to a linear trend (F (1, 7)=11.66, p=.011, η_p^2=0.625).

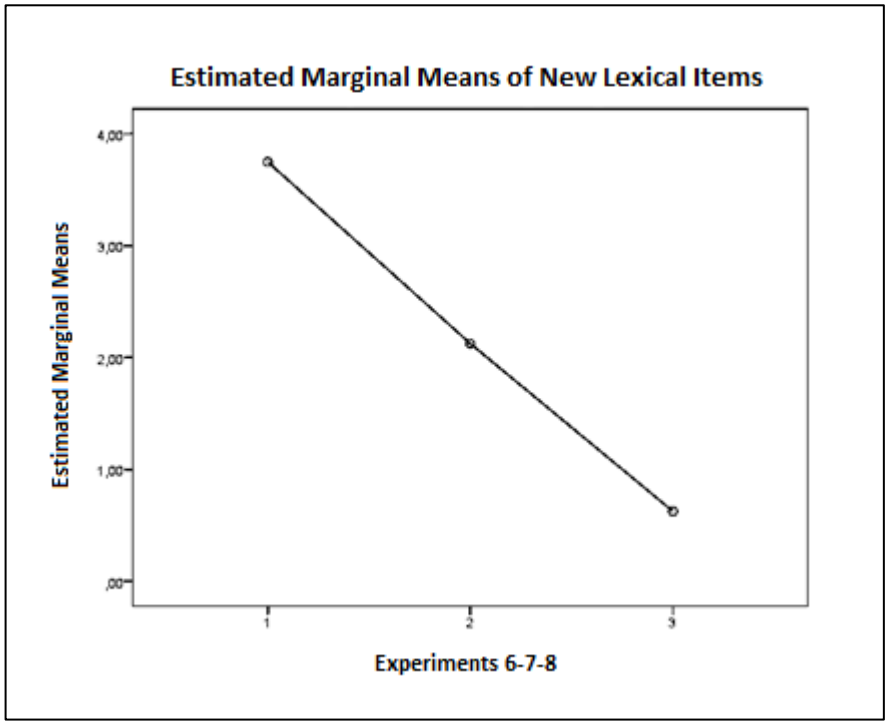


Figure 29 Estimated marginal means plot for the number of new lexical items in the Repeated Measures ANOVA test (for the 6th, 7th and 8th experiments period)

4.1.5.3. DISCUSSION

The analysis of the number of new lexical items shows a steady decrease since new lexical items are less required in the later phases thanks to the persistent use of the words created in the previous sessions. This contrasts with the situation where the lexical inventory was empty, as in the first session, or was containing only few words, as in the earlier sessions. That means, the dyad remembers and uses the previously created words when the same context or requirement is experienced. Moreover in the later sessions the need for new lexical items was limited to few

new contexts. This decrease can be considered as a criterion of saturation of the communication system in its ability to represent the experimental environment or maturation in meeting the needs of the speakers.

The decreasing trend was unaffected from the interventions at the 4th and from the break before the 7th experiment, as can be seen in Figure 30.

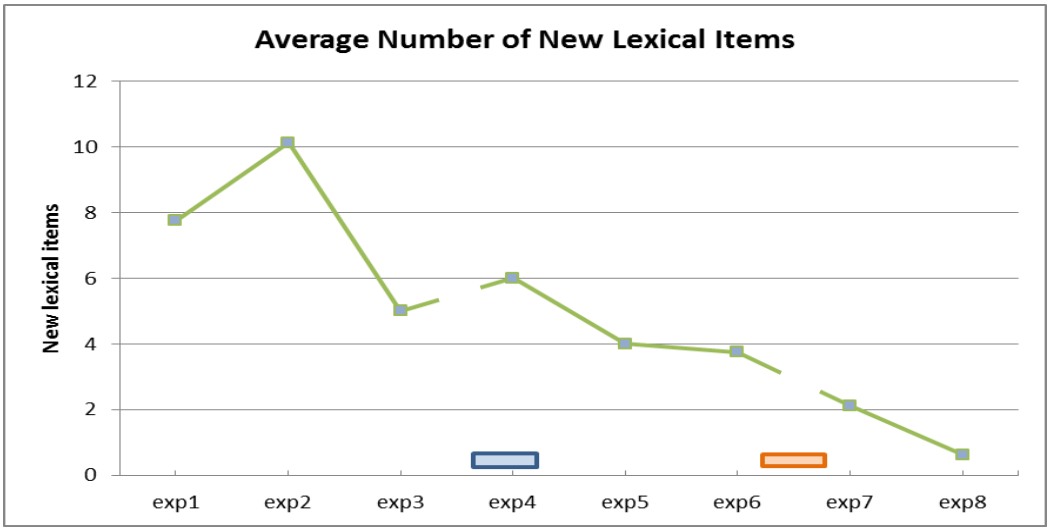


Figure 30 Average Number of new lexical items across the 8 experiments

4.1.6. TYPE NUMBER OF LEXICAL ITEMS:

The Type Number of Lexical Items means the *number of different words* used in each experimental session, where the repeated use of a frequent word is counted only once. This parameter is a measure of the richness of the lexical inventory that the participants needed to succeed in an experimental session. The type number of words can be considered as a more significant criterion for the saturation and efficiency of a communication system with respect to the token number of lexical items parameter, where unsuccessful communication and coordination may also require the use of a vast number of lexical items and bulky communication content. This parameter's trend over the eight experiments shows an increase until the mid-session phase (4-5) and then a slight decrease. This

decrease is consistent with the general trend of decrease in the turn numbers due to the efficiency increase that requires less amount of communication, however, the comparison of the decrease of the type number of words with the token number of words may be illuminating.

4.1.6.1. OVERALL CHANGE IN THE TYPE NUMBER OF LEXICAL ITEMS

The change of ***type number of lexical items*** across the eight experimental sessions was **significant** ($F(7, 49)=3.86$, $p=.002$, $\eta_p^2=0.355$). This significant result is due to a quadratic trend ($F(1, 7)=53.82$, $p<.0001$, $\eta_p^2=0.89$).

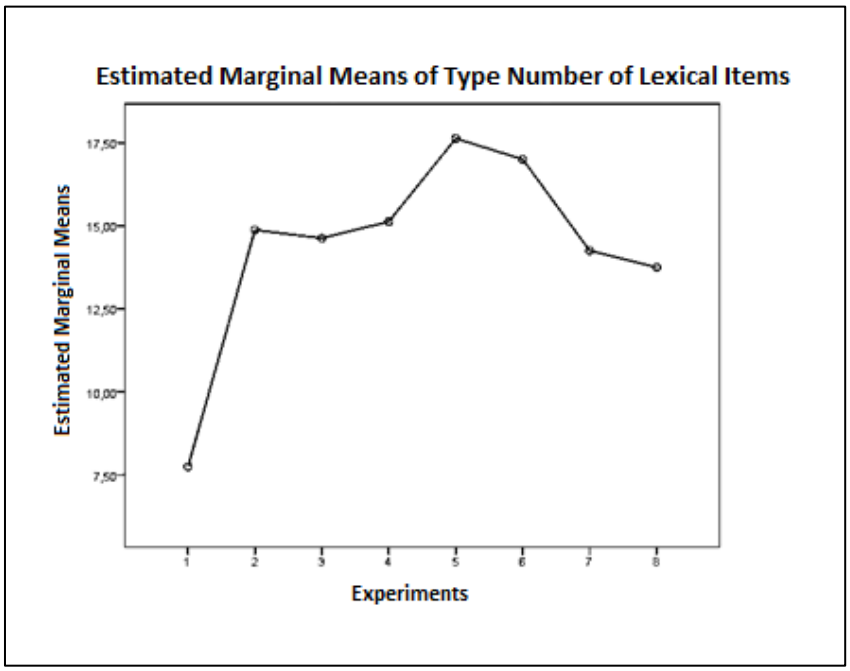


Figure 31 Average number of the type number of lexical items across 8 experiments

According to Friedman's Test, there was a statistically significant change in the type number of lexical items across the eight experiments ($\chi^2(7) = 19.980$, $p = .006$).

4.1.7. TYPE OVER TOKEN NUMBER OF LEXICAL ITEMS RATIO:

A new parameter is introduced to check the validity of the comparison of type and token number of words in each experiment and to evaluate the trend of this additional variable. The ratio is *type number of words*/*token number of words* in each experiment (henceforth "Typ/Tok")

Both parameters, in the denominator and numerator, are decreasing in the later experimental sessions but *the ratio of types over tokens is expected to increase* or at least not to decrease with the same speed with the token number of words parameter. This means that the couples speak less (more precise) with a richer lexicon (more varied). Consider, for example, couple X:

Table 18 Number of lexical types and tokens and Type/Token ratios of couple X across the 8 experiments

Couple X	∑lexical items (type) (y)	∑lexical items (token) (z)	type/token
exp1	5	17	0,29
exp2	15	26	0,58
exp3	15	23	0,65
exp4	22	88	0,25
exp5	14	41	0,34
exp6	15	57	0,26
exp7	15	37	0,41
exp8	16	24	0,66

Couple X seems to communicate more efficiently in exp8 in contrast with exp5 where there are slightly less type number of words, but the total amount of communication is much higher than it is in exp8. The dyad used 16 different words for a total of 24 times in the last experiment whereas in expe5, 14 different words are used for a total of 41 times. So this 14/41 (0.34) to 16/24 (0.66) increase in the Typ/Tok ratio is a measure of using the system more efficiently or, in other words, an increase in the ability of the system to meet the requirement of conveying the necessary information. The dyad talks less (in terms of tokens) with more words (in terms of types). This is the dictionary definition of "being concise".

The same trend as explained for Couple X, can also be seen in the other couples, see Figure 32 and Figure 33:

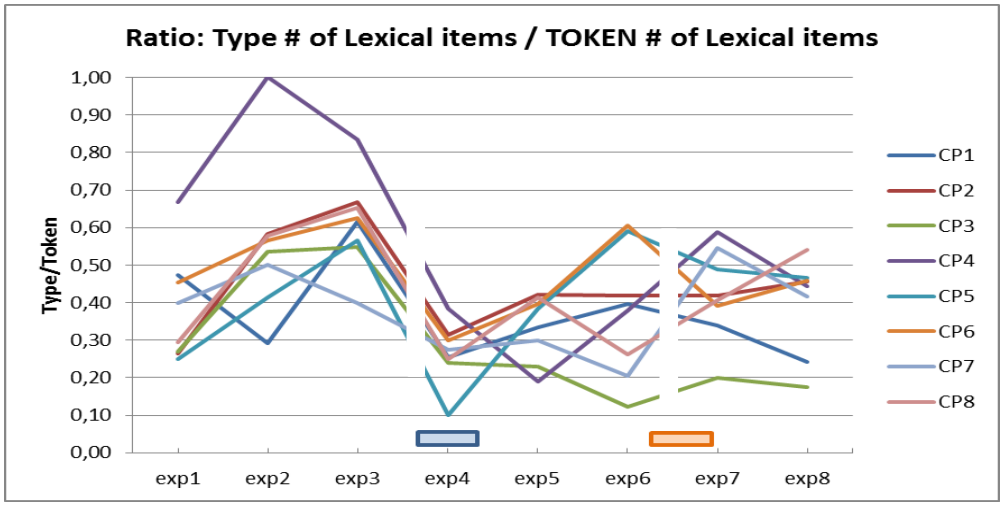


Figure 32 Average Number of new lexical items

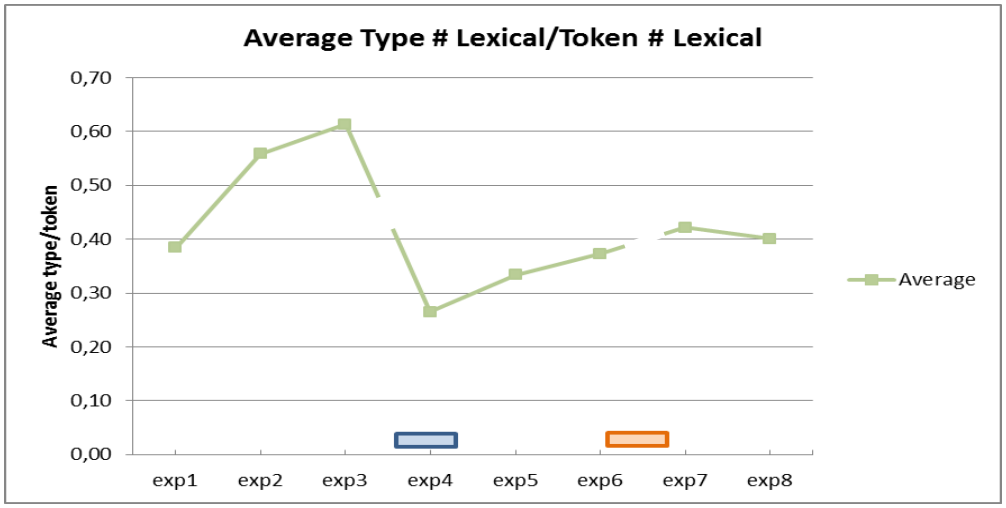


Figure 33 Average Type/Token ratio across the 8 experiments

4.1.7.1. OVERALL CHANGE IN THE TYPE OVER TOKEN NUMBER OF LEXICAL ITEMS RATIO

The change of the *type over token number of lexical items ratio* across the eight experimental sessions was **significant** ($F(7, 49)=7.29$, $p<.001$, $\eta_p^2=0.510$). This significant result is due to a cubic trend ($F(1, 7)=9.22$, $p=.019$, $\eta_p^2=0.568$).

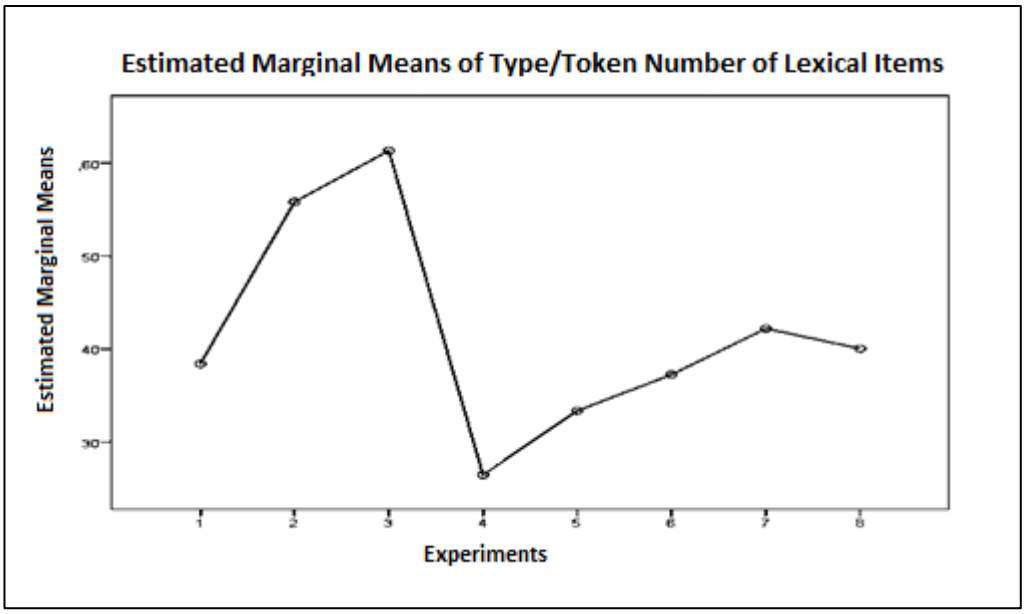


Figure 34 The estimated marginal means plot for the *type over token number of lexical items ratio*

According to Friedman's Test, there was a statistically significant change in the type/token number of words ratio across the eight experiments ($\chi^2(7) = 23.779$, $p = .001$).

4.1.7.2. THE CHANGE IN THE *TYPE OVER TOKEN NUMBER OF LEXICAL ITEMS RATIO* IN THE 4-5-6-7-8TH SESSIONS

The same Repeated Measures ANOVA test for the parameter in the two target phase of the experiment (4-5-6-7-8th experiments) also showed **a high significance** in this increasing trend. (Graph 7) The change of the ***type over token number of lexical items*** across the 4, 5, 6, 7, 8th experimental sessions was **significant** (F (4, 28)=2.74, p=.049, η_p^2=0.281). This significant result is due to a linear trend (F (1, 7)=9.72, p=.017, η_p^2=0.581).

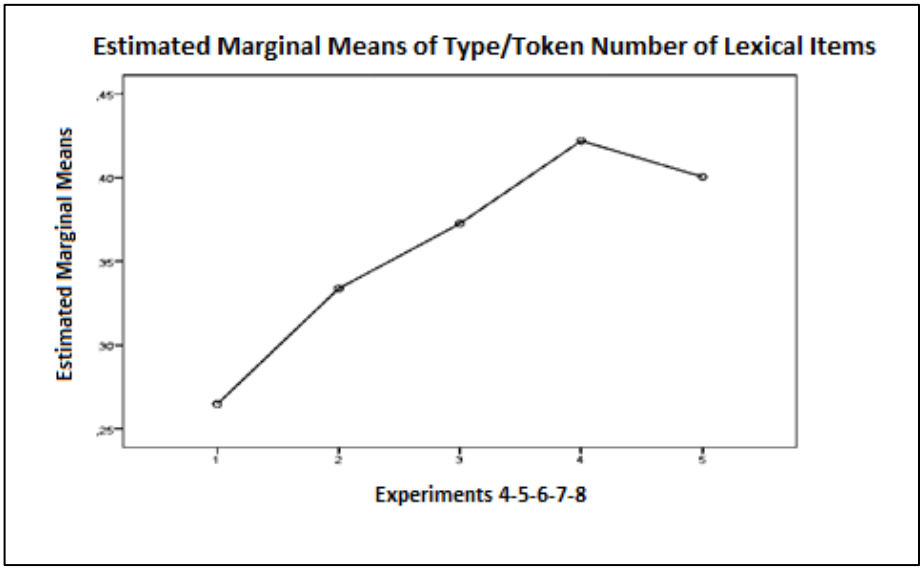


Figure 35 The estimated marginal means plot for the *type over token number of lexical items ratio* in the Repeated Measures ANOVA test (for the 4-5-6-7-8th experiments)

4.1.7.3. DISCUSSION

This trend of increase (4-5-6-7-8th) in the type/token ratio after the re-adaptation to the new task structure is strongly verified. The presentation of type/token ratio as a sound criterion for efficient communication is corroborated.

4.1.7.4. TURN TAKING RATIO P1/P2:

This parameter is analyzed to detect and measure the phenomenon of **asymmetry in turn taking** and communication between the participants within a couple. The total turn taking number of a couple may not be evenly distributed among the two participants, as two equal halves, in most of the cases. For example, in Couple 2, Participant 2 took turns eight times more than Participant 1 in the second experiment (p1/p2=0.125), but this ratio is 1 in the third experiment where both participants took equal numbers of turns. The graph in Figure 36 supports the assumption that the ratio of number of turn taking of the participants (p1/p2 or p2/p1) tends to converge on 1 on the later sessions of the experiments, mostly above a ratio of 0.80. This hypothesis claims that over time the participants of each couple tend to participate equally in the communicative collaboration, as opposed to the earlier phases of the communication system.

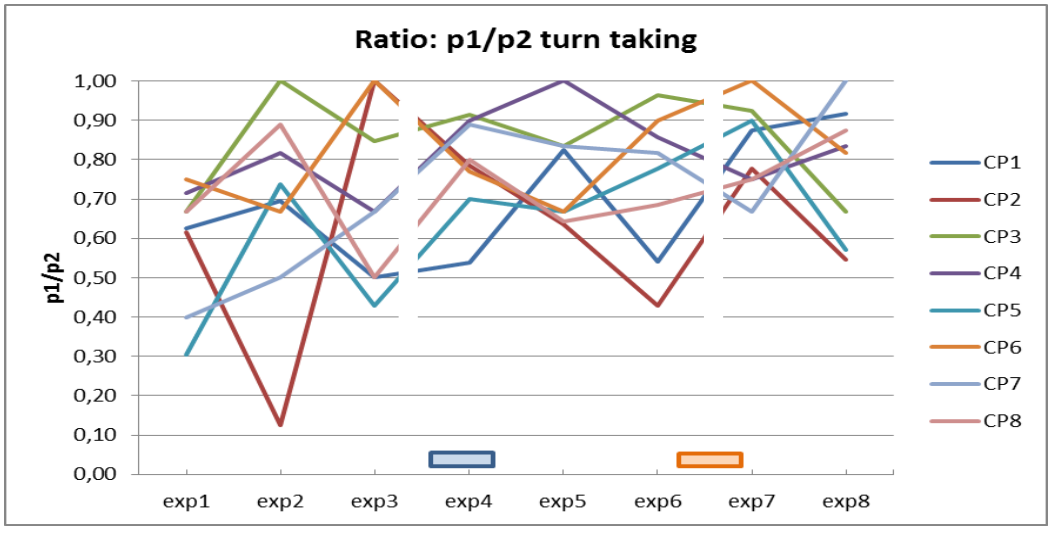


Figure 36 p1/p2 turn taking ratio of the 8 couples across the 8 experiments

Interestingly, the Repeated Measures ANOVA test *did not show a highly or even marginal significant increase* **for the ratio of turn taking numbers** over the eight experiments. (F (7, 49)=1.62, ***p=.153***, η_p^2=0.188). Although a positive trend can be seen in Figure 37, the variation is too high to make this trend significant.

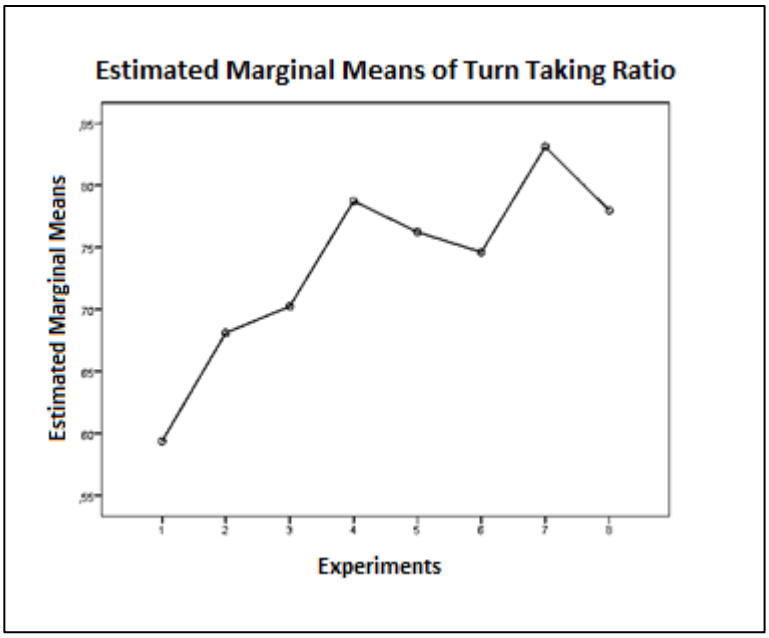


Figure 37 The estimated marginal means plot for the *turn taking ratio* in the Repeated Measures ANOVA test (for the eight experiments)

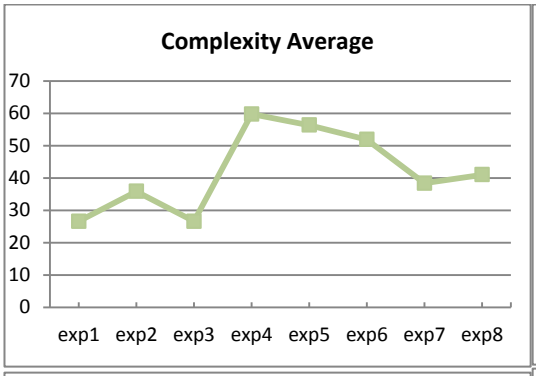
4.2. SYNTACTIC COMPLEXITY ANALYSIS:

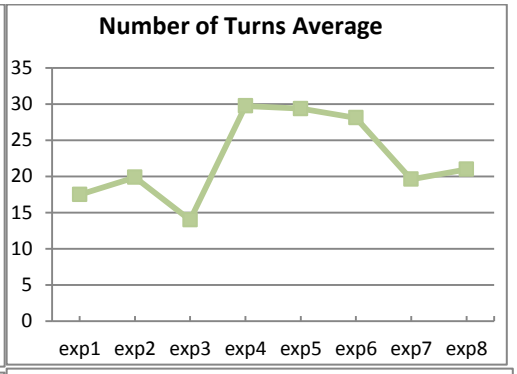
The syntactic complexity parameter is used to test the change or difference in syntactic complexity of the utterances across the experimental sessions. There are three ordinal categories, 1, 2 or 3, which are assigned to expressions of the participants, respectively, and whose averages quantify the total complexity in each experiment.

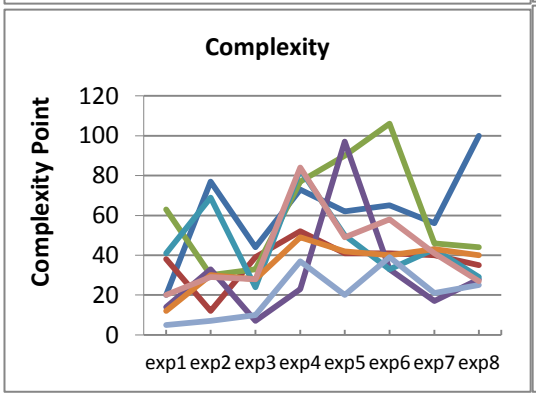
4.2.1. COMPLEXITY PARAMETER**:**

This parameter's trend over the eight experimental sessions shows some variance among couples and during the different phases of the experiments, as can be seen in Figure 38. Interestingly, the observed trend is very parallel with the trends

of turn numbers which means that each couples' average complexity point for a turn is very close to the number of turns, as can be seen by comparing the complexity average in Figure 38 and the number of turns average in Figure 39 . Thus, we can predict the complexity point for an experiment from the turn numbers in that experiment. This finding requires investigating whether syntactic complexity is a redundant variable.







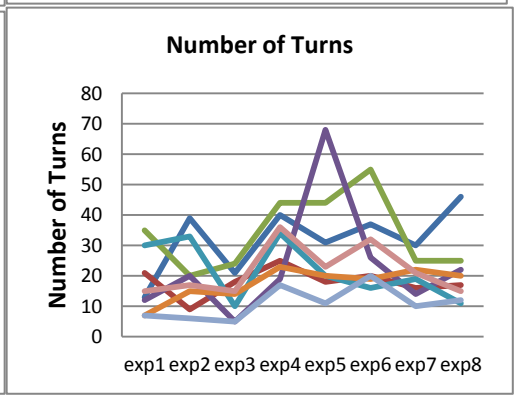


Figure 38 Complexity

Figure 39 Number of Turns

To filter out the effect of inter-couples differences about number of turns for each experiment over the complexity scores, the complexity score is normalized by using the complexity score / number of turns ratio for each experiment. By doing this we can remove the effect of the increase or the decrease in the amount of the communication content. The complexity score is the total of the complexity scores of turns in an experimental session. If the session is a longer one (comprises of more number of turns), total score of complexity will be higher, independent from

the average complexity of the turns. Hence, using the *complexity score / number of turns ratio* filters out the effect of length of session (in terms of number of turns) over complexity score for that session. This yields the following trend, in

Figure 40, where the average complexity score across the 8 experiments **still shows a slight increase** from a point where the utterances usually have type2 and type1 complexity to a point where the complexity is somewhere between type3 and type2 utterances.

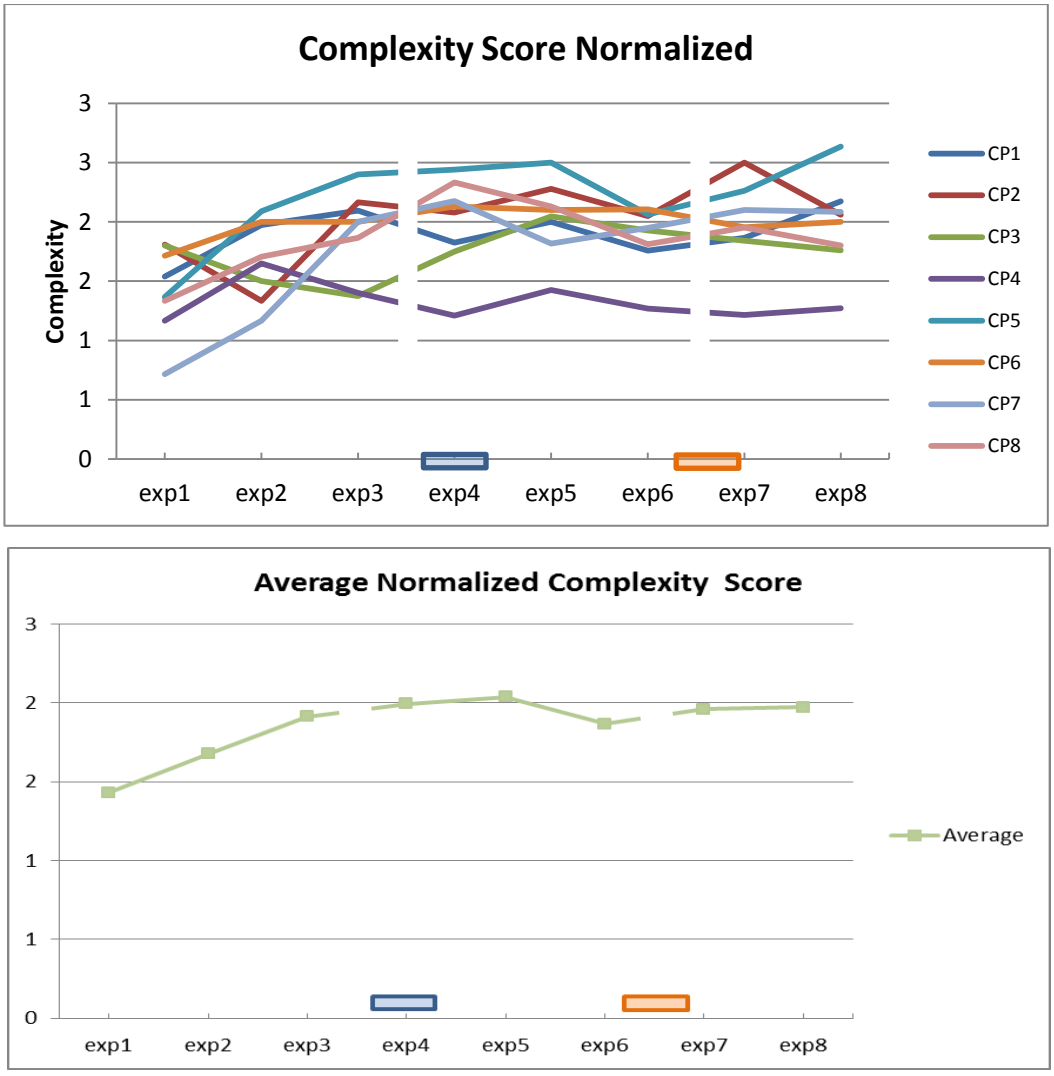


Figure 40 Normalized (upper panel) and average normalized (lower panel) complexity scores for the 8 couples across 8 experiments

When the change of the average complexity score is compared using Figure 38 and Figure 40, we can see that normalization filters out the effect of the increase in total turn numbers in the 4th experiment and the slight increase of complexity is unaffected from the change that causes a rise in the total amount of communication.

The change of **the complexity score** (in the non-normalized data) across the eight experimental sessions was **significant** under the Greenhouse-Geisser correction ($F(3.5, 24.54) = 4.35$, $p=.011$, $\eta_p^2=0.383$). This significant result is due to a significant linear trend ($F(1, 7) = 6.57$, $p=.037$, $\eta_p^2=0.484$), a quadratic trend ($F(1, 7)=21.91$, $p=.002$, $\eta_p^2=0.758$) and a significant quartic (fourth order) polynomial trend ($F(1, 7)=8.85$, $p=.021$, $\eta_p^2=0.56$)

According to Friedman's Test, there was a statistically significant change in the syntactic complexity scores across the eight experiments ($\chi^2(7) = 27.854$, $p < .0001$).

The normalized data has a similar trend of increase. The change of **the normalized complexity score** across the eight experimental sessions was **significant** under the Greenhouse-Geisser correction ($F(2.59, 18.16) = 5.35$, $p=.010$, $\eta_p^2=0.433$). This significant result is due to a significant linear trend ($F(1, 7) = 9.42$, $p=.018$, $\eta_p^2=0.57$) and a quadratic trend ($F(1, 7)=12.77$, $p=.009$, $\eta_p^2=0.65$)

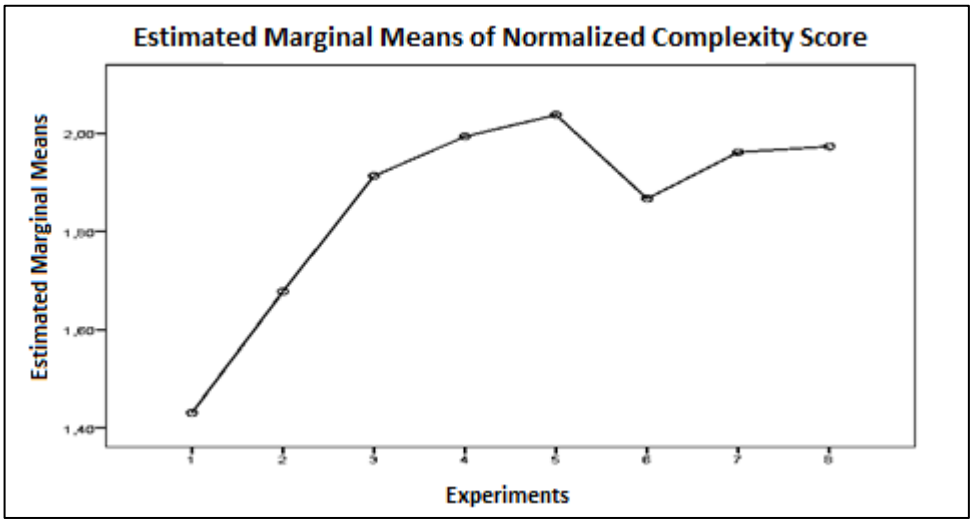


Figure 41 Estimated marginal means plot for the normalized complexity score in the Repeated Measures ANOVA test

According to Friedman's Test, there was a statistically significant change in the normalized syntactic complexity scores across the eight experiments ($\chi^2(7) = 19.997$, $p = .006$).

4.3. LEXICAL CATEGORY ANALYSIS

After the quantitative variables, the communication content is categorized into lexical items. Lexical items are identified, counted and categorized in terms of classical grammatical categories: Noun, Pronoun, Verb, Adjective/Adverb, Preposition, Interjection, Conjunction, Symbol.

Note: "*symbol*" is the lexical category where the two symbols "→" and "**?**" provided to the participants are used for their graphic properties. However, cases like → for "follow" are not counted as "symbol" but as an iconic sign.

4.3.1. TOKEN NUMBER OF NOUNS

The number of nouns is used as the first and most important lexical category parameter. However when the change of this token count is visualized, the change across the experiments and the variation among the couples seems to be seriously affected by the amount of communication in certain experimental sessions. In the following Figure 42 the high numbers of nouns are due to the general amount of increased communication during **exp5** for **Couple 4** (purple line), and **exp8** for **Couple 1** (dark blue line).

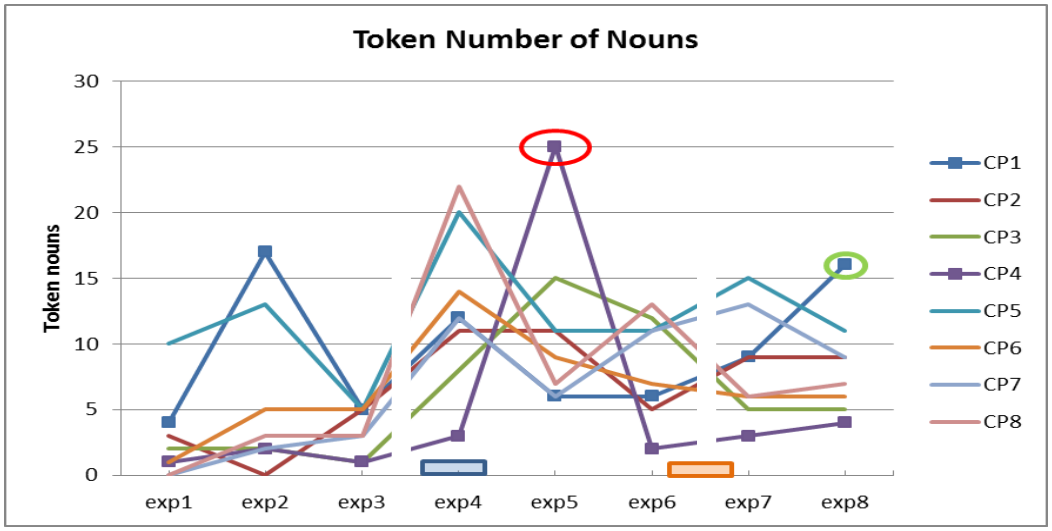


Figure 42 Token number of nouns (CP1 is Couple 1, CP2, … CP8)

The number of use of the other lexical categories is higher for the specific couples and experimental sessions. But to see the effect of total amount of communication, the normalization procedure is used which is dividing the number of lexical categories by the total turn number of lexical items in an experimental session. The normalized data is shown in Figure 43.

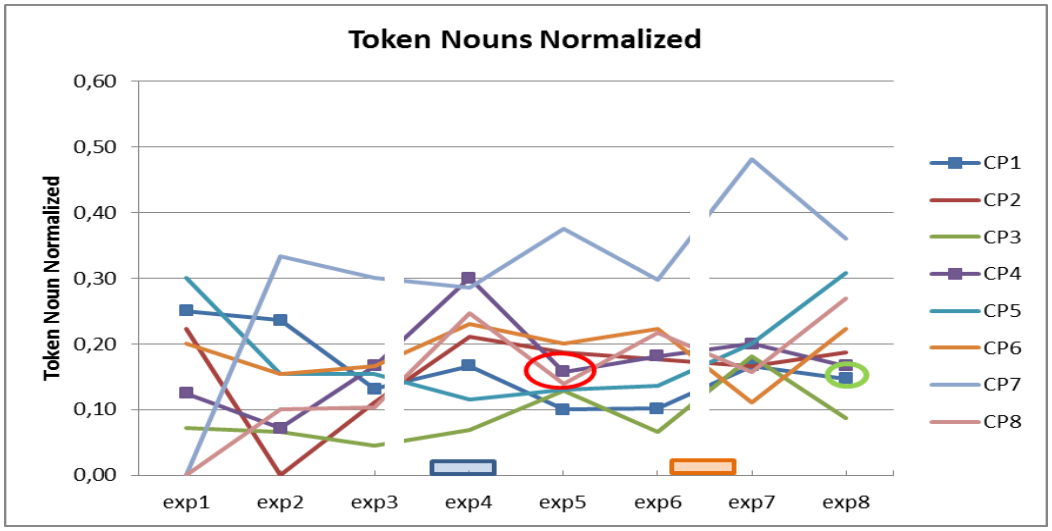


Figure 43 Token Nouns Normalized

As it can be seen from the data of **exp5** for **Couple 4** (purple line) in the normalized data (Figure 43), unlike in the Figure 42, the rate of use of nouns is not higher than that of the other couples. Couple 4 used all the lexical categories a lot in the 5th session. There is nothing special about Couple 4 and their usage of nouns. This example shows that the information gained from the normalization procedure is quite different from the information we get just with the token number of nouns. To filter out the effects of other variables this "normalization" procedure is applied **to the other token counts of lexical categories** as well.

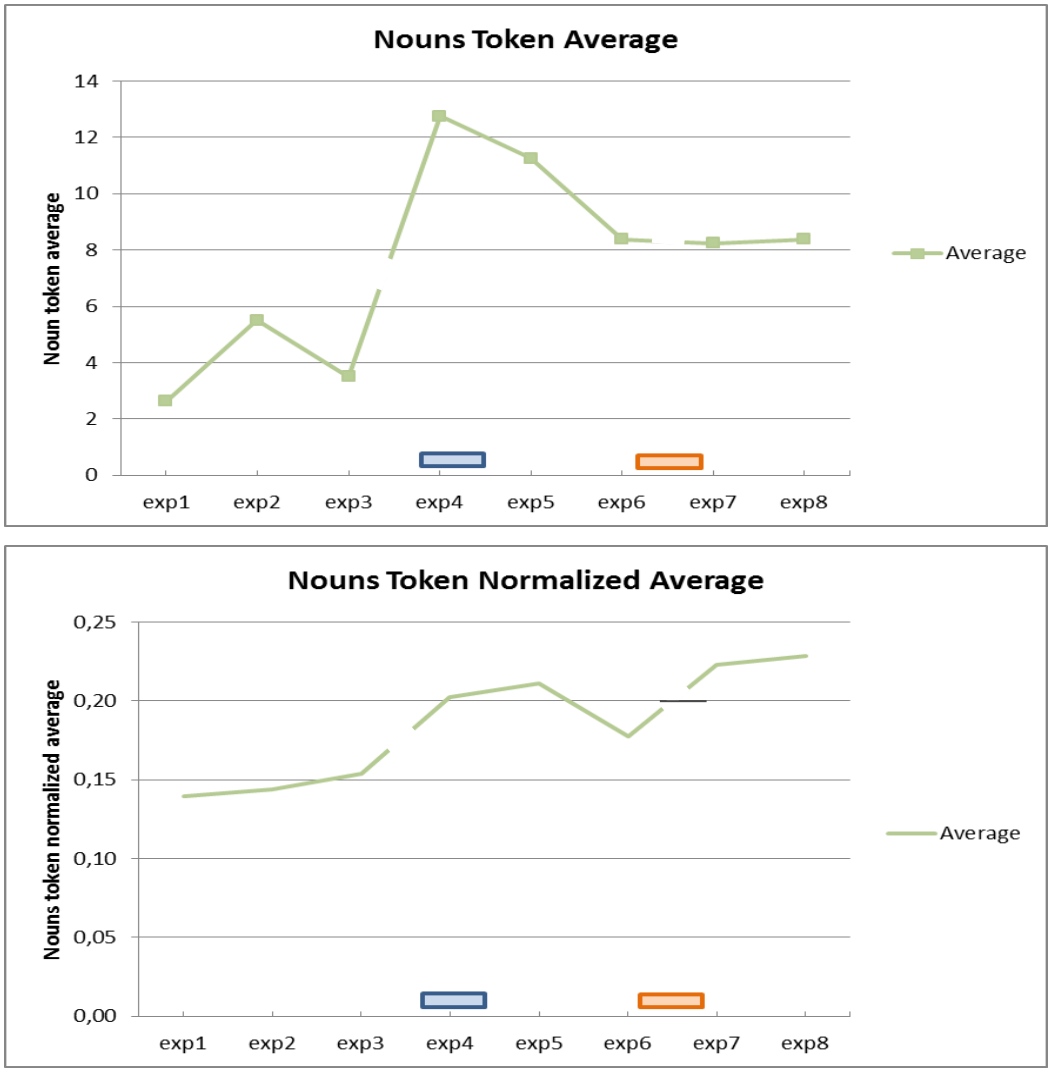


Figure 44 Token Nouns Average / Nouns Normalized Average

The normalization provides a better visual representation of the ratio of nouns to the other lexical categories. For example the graphs of average token number and normalized token number shows the same increase, but the normalized graph describes a smoother increase during exp3 and exp4 and exp5.

This increase is statistically tested by Repeated Measures ANOVA. The change of token number of nouns across the eight experimental sessions was **significant** ($F(7, 49)=5.05$, $p<.001$, $\eta_p^2=0.419$).

According to Friedman's Test, there was a statistically significant change in the token number of nouns across the eight experiments ($\chi^2(7) = 35.620$, $p < .0001$).

This significant result is due to a linear trend (F (1, 7)=29.09, p<.000, η_p^2=0.81), a quadratic trend (F (1, 7)=10.68, p=.014, η_p^2=0.60) and a significant quartic (fourth order) polynomial trend (F (1, 7)=5.58, p=.05, η_p^2=0.44).

A Repeated Measure ANOVA on the **normalized data** across the 8 experiments yielded **no significant effect** (F (7, 49)=1.76, p<.117, η_p^2=0.201).

According to Friedman's Test, the change in the normalized number of nouns across the eight experiments **was not significant** ($\chi^2(7) = 11.555$, $p = .116$).

4.3.2. TOKEN NUMBER OF PRONOUNS

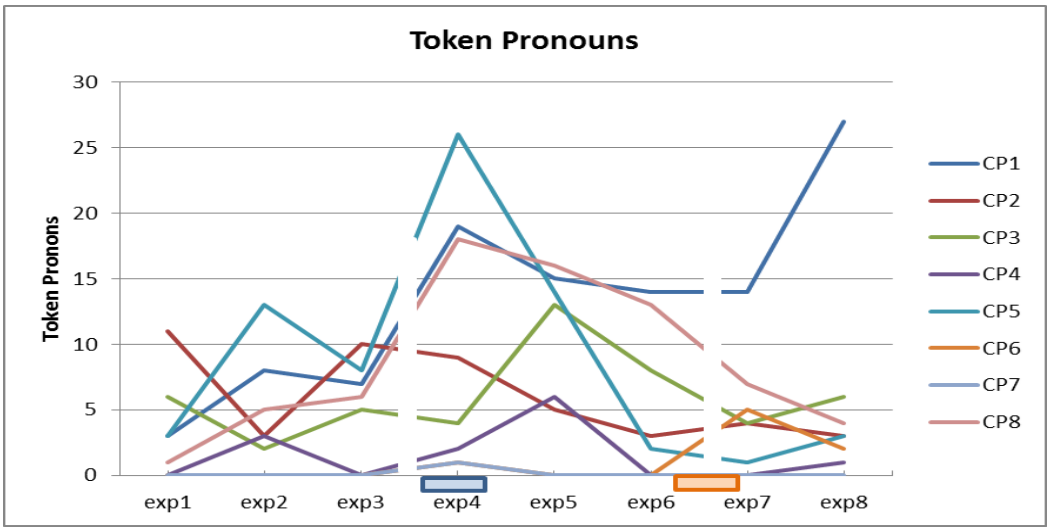


Figure 45 Average Number of Pronouns

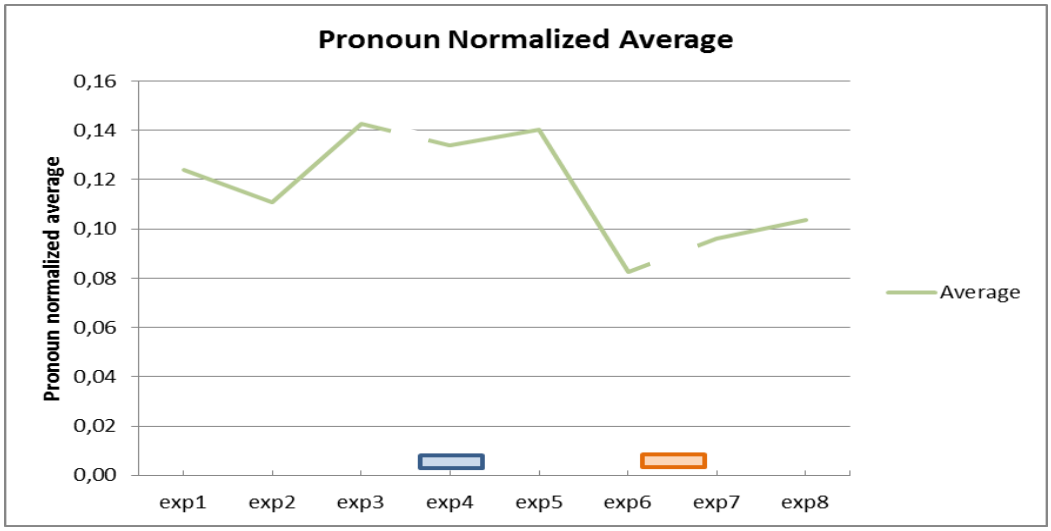


Figure 46 Average Normalized Counts of Pronouns (use ratio)

The token count of pronouns show high variation among couples and across the eight experiments and appears to be affected by the overall amount of communication in each session. The normalization shows that the increase in numbers during the critical exp3-exp4 phase does not change the proportion of usage of pronouns with respect to the other lexical categories. But there is a trend of slight decrease through the 8 experiments and a more visible decrease in the exp4-5-6-7-8 series.

The Repeated Measures ANOVA on **the token** and **normalized token number of pronouns** shows no indication of a significant decrease for both of the data descriptions. Pronouns Normalized: The change of the normalized number of pronouns across the eight experimental sessions was **not significant**. ($F(7, 49)=0.85$, $p=.551$, $\eta_p^2=0.108$).

According to Friedman's Test, the change in the normalized number of pronouns across the eight experiments **was not significant** ($\chi^2(7) = 6.976$, $p= .431$).

Pronoun Token: The change of token number of pronouns across the eight experimental sessions was **not significant**. (F (7, 49)=2.10, p<.131, η_p^2=0.231).

According to Friedman's Test, the change in the token count of pronouns across the eight experiments **was not significant** ($\chi^2(7) = 13.467$, $p = .062$).

The Repeated Measures ANOVA on **the token** and **normalized token number of pronouns** for the exp4-5-6-7-8 series shows no indication of a significant decrease for both of the data descriptions.

4.3.3. TOKEN NUMBER OF VERBS:

The token and normalized token number of verbs shows a general increasing trend in the earlier experimental sessions and then follows a steady trend throughout the remaining sessions.

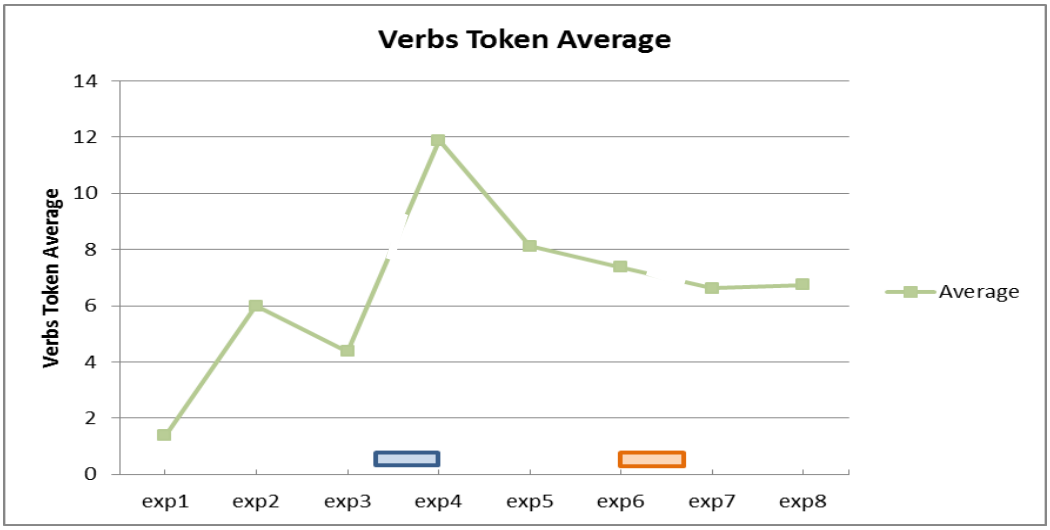


Figure 47 Average Number of Verbs across the 8 experiments

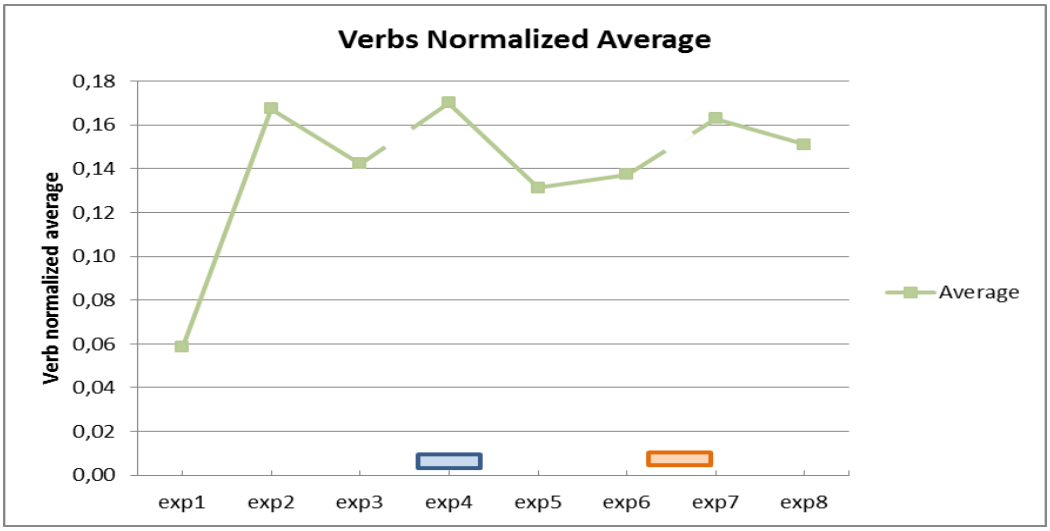


Figure 48 Average Normalized Counts of Verbs (use ratio) across the 8 experiments

The change of **token number of verbs** across the eight experimental sessions was **significant** under the Greenhouse-Geisser correction ($F(2.99, 20.98) = 4.38$, $p=.015$, $\eta_p^2=0.385$). This significant result is due to a linear trend ($F(1, 7) = 5.71$, $p=.048$, $\eta_p^2=0.45$), and a quadratic trend ($F(1, 7)=5.89$, $p=.046$, $\eta_p^2=0.46$).

According to Friedman's Test, there was a statistically significant change in the token number of verbs across the eight experiments ($\chi^2(7) = 28.070$, $p < .0001$).

The change of **normalized number (ratio) of verbs** across the eight experimental sessions was **significant** ($F(7, 49) = 3.94$, $p=.002$, $\eta_p^2=0.360$). This significant result is due to a cubic trend ($F(1, 7) = 9.99$, $p=.016$, $\eta_p^2=0.588$).

According to Friedman's Test, there was a statistically significant change in the normalized number of verbs across the eight experiments ($\chi^2(7) = 17.950$, $p= .012$).

4.3.4. TOKEN NUMBER OF ADJECTIVES AND ADVERBS

The two categories of adjectives and adverbs are counted separately but unified in the descriptive data analysis phase from the very earlier stages of the study. The token numbers of this combined category shows an increase in the

earlier sessions of the experimental sessions and then a slight decrease, as can be seen in Figure 49. The normalized data shows an early increase and then a steady trend, as can be seen in Figure 50.

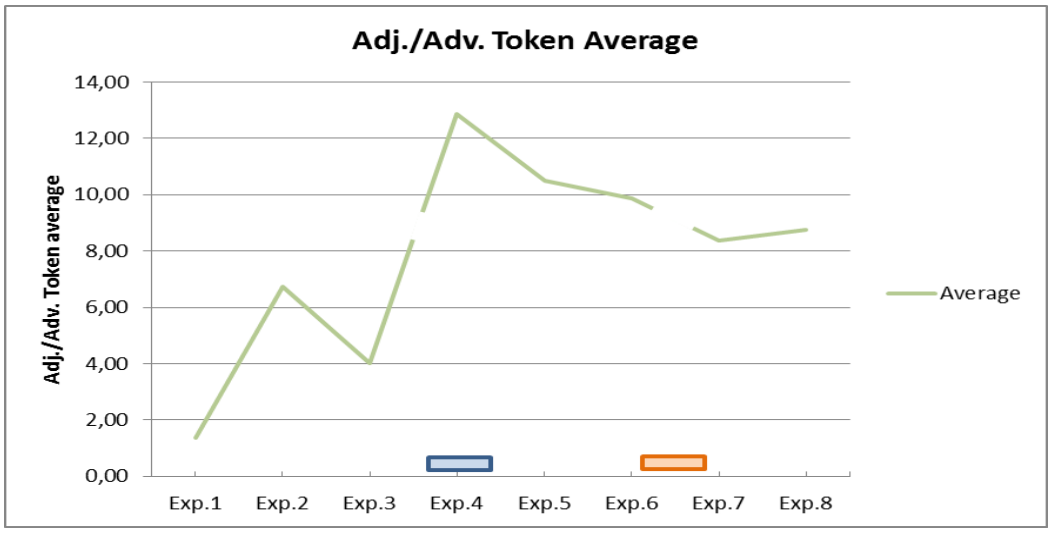


Figure 49 Average Number of Adj./Adv. Across the 8 experiments

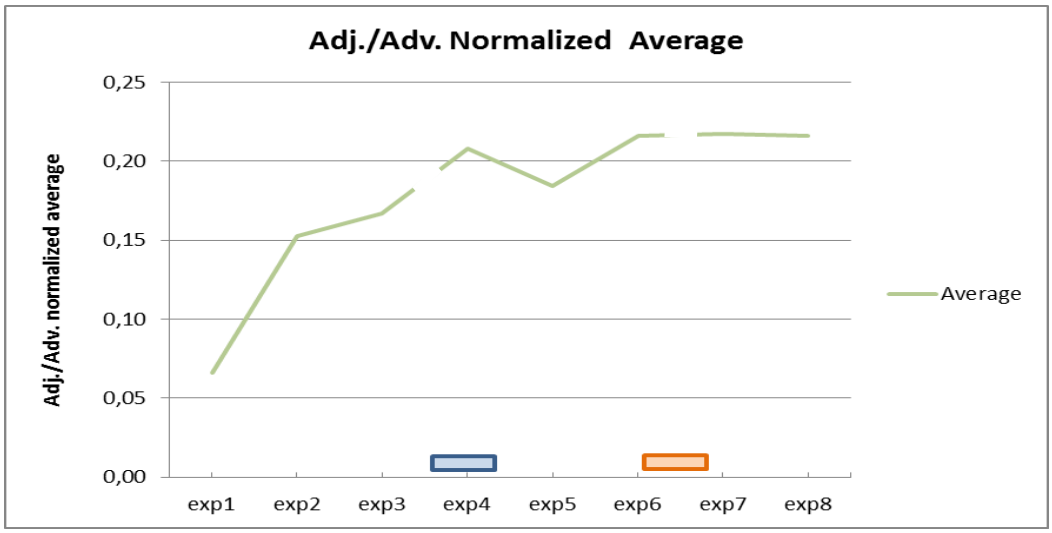


Figure 50 Average Normalized Adj./Adv. (Use ratio) across the 8 experiments

The increase from the first to the eighth experiment shows a highly significant increase as revealed by a Repeated Measures ANOVA for both the **token number of Adj./Adv.** and **normalized token number of Adj./Adv.** data sets.

The change of **token number of Adj./Adv.** across the eight experimental sessions was **significant** ($F(7, 49)=5.36$, $p<.001$, $\eta_p^2=0.434$). This significant result is due to a linear trend ($F(1, 7)=17.22$, $p=.004$, $\eta_p^2=0.71$) and a quadratic trend ($F(1, 7)=19.27$, $p=.003$, $\eta_p^2=0.734$).

According to Friedman's Test, there was a statistically significant change in the token number of Adj./Adv. across the eight experiments ($\chi^2(7) = 27.514$, $p < .0001$).

The change of the **normalized number of Adj./Adv.** across the eight experimental sessions was **significant** under the Greenhouse-Geisser correction ($F(2.95, 20.62)=3.51$, $p=.034$, $\eta_p^2=0.334$). This significant result is due to a significant linear trend ($F(1, 7)=9.01$, $p=.02$, $\eta_p^2=0.563$).

4.3.5. TOKEN NUMBER OF INTERJECTIONS

Interjections are words used as exclamations, expressions of emotions and sentiment. Words like "hi", "ok", "well", "yes", "no" are also categorized as interjections.

Interjections are very abundant in the earlier sessions of experiments due to their function in speech acts which contain directives and expressive where participants need to check the attention, availability and the comprehension of the partner.

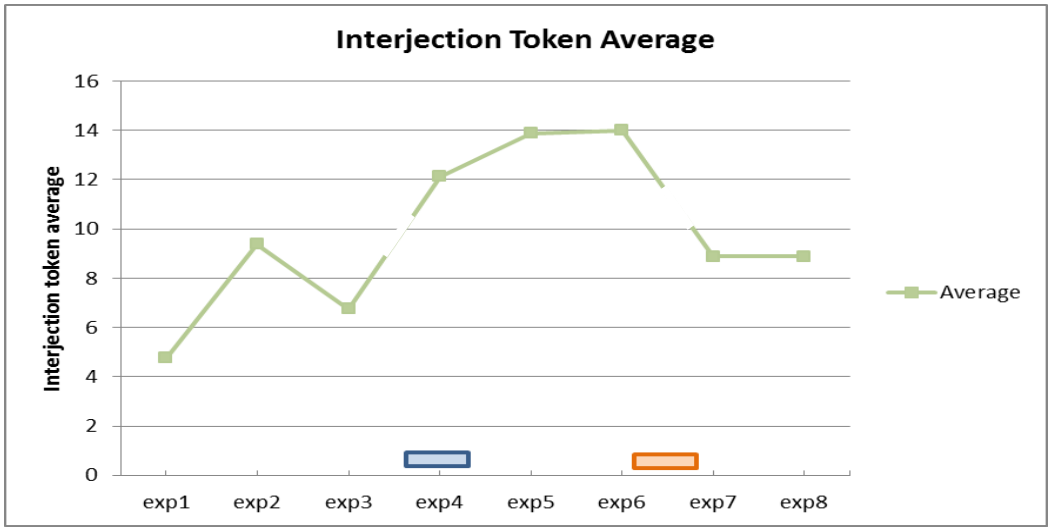


Figure 51 Average Token Number of Interjections across the 8 experiments

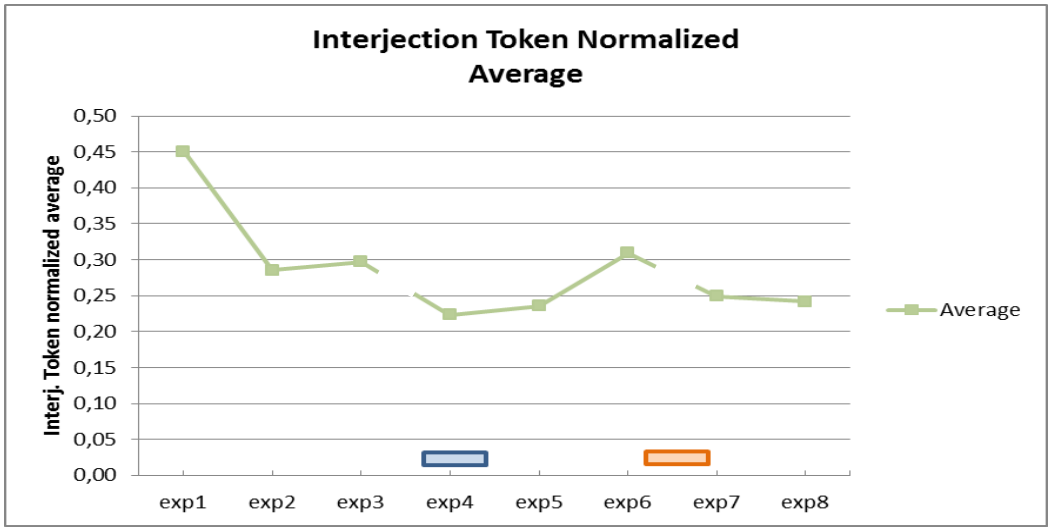


Figure 52 Average Normalized Number of Interjections across the 8 experiments

As can be inferred from the normalized data in the earlier sessions, nearly half of the lexical items – out of a total of nine lexical categories – were interjections. Considering that the later proportional decrease that can be seen in the normalized data (see Figure 52), and considering the increase in the total number of lexical items in the later sessions the increase in token count of interjections does not

differ from the increase of the remaining categories. Even more, the normalized data show a relative decrease with respect to the other variables.

The change of **token number of interjections** as revealed by a The Repeated Measures ANOVA across the 8 experiments was **significant** (F (7, 49)=4.34, p=.001, η_p^2=0.383). This significant result is due to a linear trend (F (1, 7)=12.04, p=.001, η_p^2=0.632), a quadratic trend (F (1, 7)=12.08, p=.003, η_p^2=0.633), and a fifth order polynomial trend (F (1, 7)=7.06, p=.033, η_p^2=0.502).

According to Friedman's Test, there was a statistically significant change in the token number of interjections across the eight experiments ($\chi^2(7) = 24.865$, p = .001).

The more important assumption concerns the normalized data which assumes a relative decrease for the normalized number (ratio of) interjections; however, a Repeated Measures ANOVA revealed that the normalized number of interjections **did not change significantly** over the course of the eight experiments under the Greenhouse-Geisser correction (F (1.66, 11.61)=2.716, *p=.113*, η_p^2=0.280), although numerically there is an increase, as can be seen in Figure 53 below:

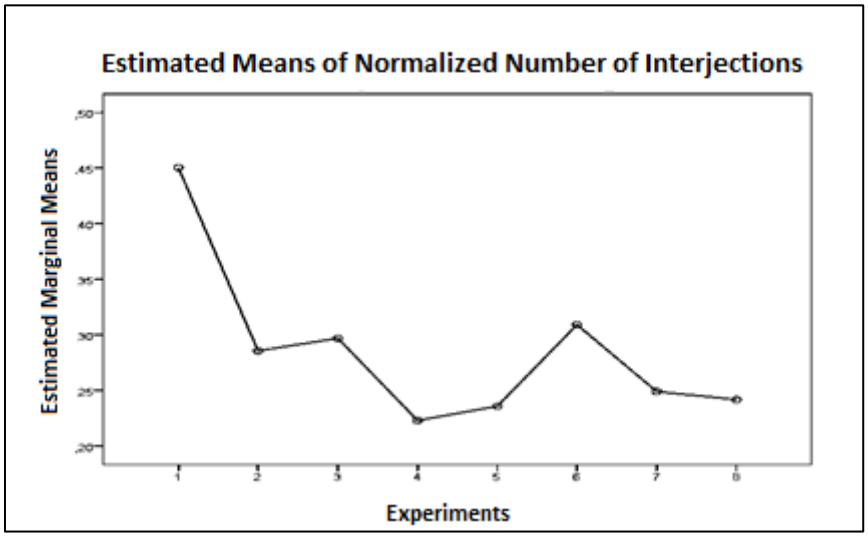


Figure 53 Estimated Means of Normalized Number of Interjections

According to Friedman's Test, the change in the normalized number of interjections across the eight experiments **was not significant** ($\chi^2(7) = 9.540$, $p = .216$).

4.3.6. THE OVERALL ANALYSIS OF LEXICAL CATEGORIES

The Combined graphs of the average token count and normalized data for all five major lexical categories are as follows (the very rarely used categories symbol, conjunction and preposition is not included here):

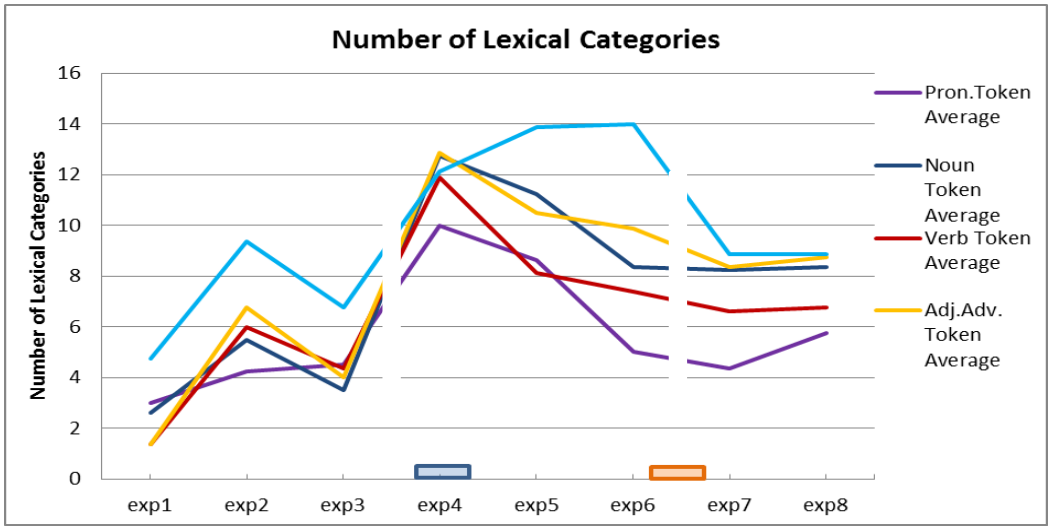


Figure 54 The average token number of lexical categories across the 8 experiments

Note that the graph for the Adjective/Adverb category is very parallel to that of Nouns. The Interjections are the highest in the beginning and following sessions; however, decrease in the later sessions.

The data without interjections (see Figure 55 below) shows the general parallel trend for the other lexical categories despite the superiority of nouns and adjectives over the remaining ones. Interjections are excluded in order to change the scale and make the trends of other variables more visible.

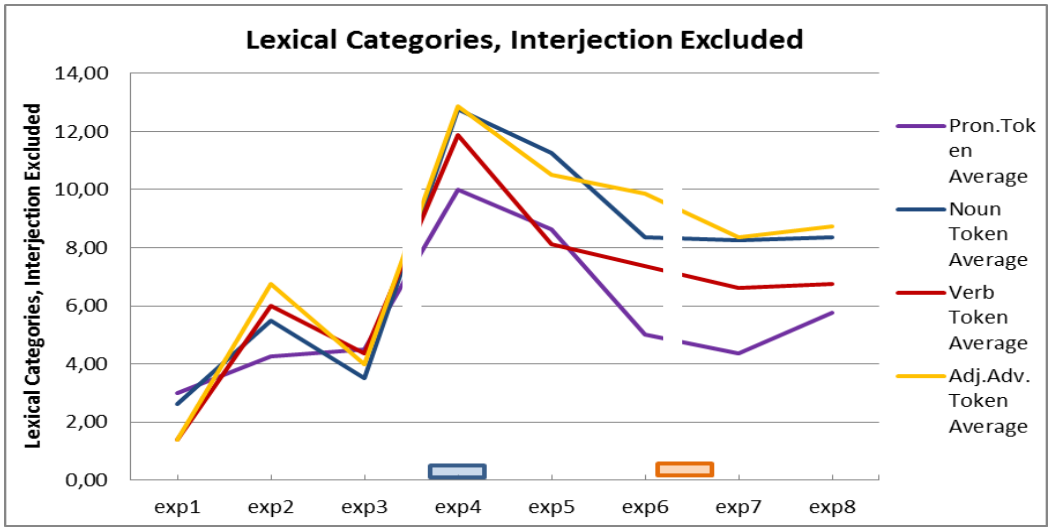


Figure 55 Average token number of lexical categories across the 8 experiments, interjection excluded

The normalized data visualizes some different trends when the increase, or decrease in, use ratios of the various lexical categories are compared:

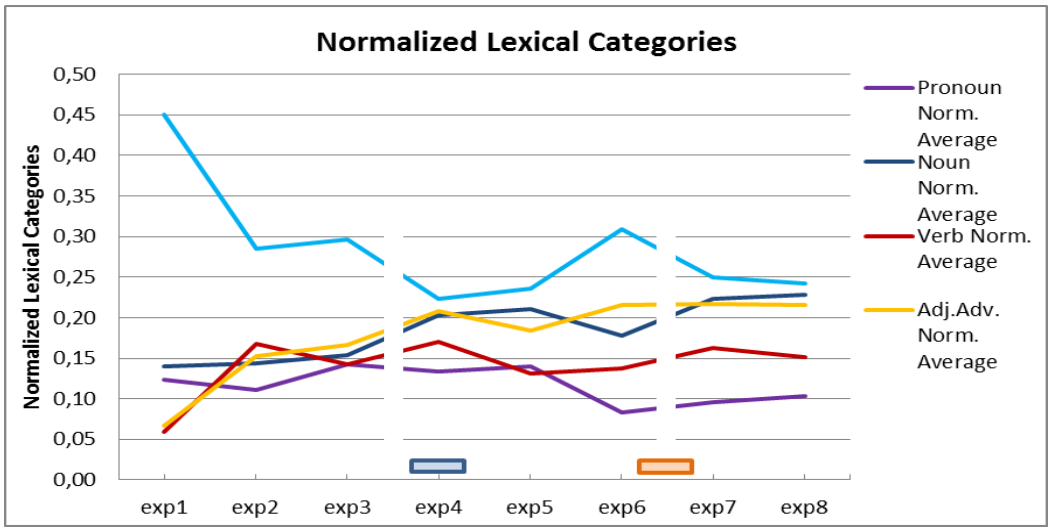


Figure 56 The average normalized number (use ratio) of lexical categories across the 8 experiments

The usage ratios over total turns (normalized data) in each experiment depicts a decreasing trend for interjections and a weaker one for pronouns; verbs are stable and adjective/adverb and noun categories show a slight increase in the later sessions (see Figure 56).

The following graph (Figure 57) illustrates the trends of the prototypical categories nouns, pronouns, verbs and adjectives/adverbs, excluding interjections. (Interjections are excluded to change the scale and make the trends of the other variables more visible):

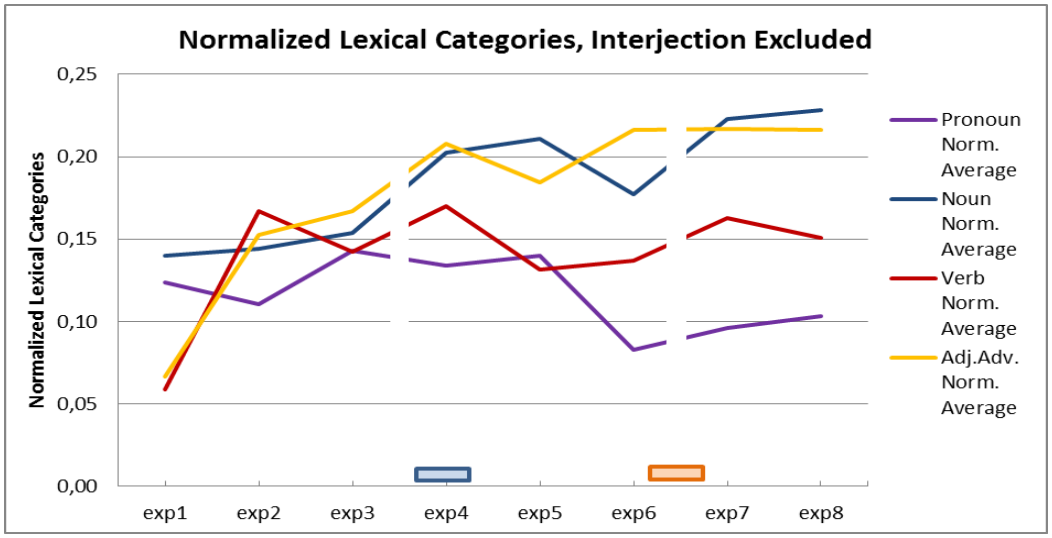


Figure 57 The average normalized number (use ratio) of lexical categories across the 8 experiments, interjections excluded

The parallel increasing trend of nouns and adjectives and the slight decrease in pronouns is more visible in. Even though we can see a decreasing trend for all categories in the “token numbers” graph, this decrease is about the increase of the efficiency of the communication system and the use of it by the participants. That is, the decrease is about the general decrease in the amount of communication after the mid-sessions. However, when the amount of communication is filtered by

normalizing the token numbers by the total number of lexical items we see an opposing trend of increasing (not decreasing) use of nouns (and a parallel trend of adjectives) and a decreasing trend of pronouns and interjections.

In terms of total number of uses (tokens), all categories such as **noun**, **verb**, **adjective/adverb**, and **interjection,** we found a statistically significant change across the eight experiments. But since this change may be claimed to be due to the total amount of communication, there was a need for normalization of this token count of categories by the total number of them. As for the normalized data of lexical categories and trends referring to them, the changes were **not significant** for the **nouns, pronouns** and **interjections**. The change for **adjective/adverbs** and **verbs** was **significant**. (For the **pronouns** the change was not significant for both the normalized and token count data).

The numerical decrease of interjections in their normalized use ratios and the relative decrease in the token count shows a trend towards more complex and NL-like utterances instead of mere interjections, attention-grabbing, availability-checking, comprehension- checking, exclamations, such as “??”, “oq?”, “ye?”, “oey”, “yeee”, “y” etc.). Interjections are used when the participants need to make more confirmations, clarifications and when they need to correct misunderstandings. When the experience and skill and the communication system develops there are less numbers of similar lexical items needed for such purposes.
Adjectives/adverbs and nouns on the contrary are the carriers of objective information about the external world, navigation and the task content. The couples are observed to use them more often relatively to the other categories in all of the stages of the experiments – earlier, mid or later sessions (see Figure 56 and Figure 57 above). Their relative use, (use ratio) did not “suffer” from the decrease of amount of communication in the later sessions. In the 3rd experiment where most of the couples figured out how to create and use their communication system, the average use ratio of Interjections was 30 % and the average use ratio of other lexical

categories (Adj/Adv, Noun, Pronoun, Verb) was around 15 %. In the later sessions, it is around 25 % for Interjections, Nouns and Adj/Adv. categories and same 15% for Verbs and 10% for Pronouns (see Figure 56 and Figure 57). This relative decrease in the use ratio of pronouns can be related to less need for navigational coordination and for organizing/regulating/reassuring correct communication by communicative action. Pronouns are usually used in “Where are you?”, “Are there?”, “Are you ok?”, “Did you understand me?”, “What is it?”-like sentences and in the repetition of “I am there”, “You follow me”, “I follow you”-style sentences which are both more needed in the earlier sessions of less efficient adaptation to the environment and task and less effective communication systems.

4.4. SPEECH ACT CATEGORY ANALYSIS

Each turn taking (utterance) of the participants is categorized into a speech act. Experiment-wise data are collected by counting the speech acts, similar to the other analysis levels. These data are summarized into a table in terms of totals and normalized by dividing each experimental session’s category totals by the total number of turns in that experiment.

The categories seem to be intersecting and equivocal. Referring to the literature about speech act categories, a set of more certain instructions was prepared for categorization. A second rater repeated the whole categorization procedure. A Kappa Inter-rater agreement test was done to check the used categories’ appropriateness to measure or characterize speech acts. Note that 100 % of the data of Speech Act categorization was rated by the second rater. The results of separate tests run for each couples’ data is as follows:

CP1: The inter-rater reliability for the raters was found to be: Kappa = 0.82 ($p < .0.001$)
CP2: The inter-rater reliability for the raters was found to be: Kappa = 0.79 ($p < .0.001$)

CP3: The inter-rater reliability for the raters was found to be: Kappa = 0.85 (p <.0.001)
CP4: The inter-rater reliability for the raters was found to be: Kappa = 0.68 (p <.0.001)
CP5: The inter-rater reliability for the raters was found to be: Kappa = 0.87 (p <.0.001)
CP6: The inter-rater reliability for the raters was found to be: Kappa = 0.89 (p <.0.001)
CP7: The inter-rater reliability for the raters was found to be: Kappa = 0.89 (p <.0.001)
CP8: The inter-rater reliability for the raters was found to be: Kappa = 0.73 (p <.0.001) Average Kappa = 0.83

The above available Kappa scores revealed the raters' reliability were in substantial agreement.

4.4.1. ASSERTIVE SPEECH ACTS:

This category of speech act is counted both as tokens and normalized. The trends are depicted in Figure 58 and Figure 59 below:

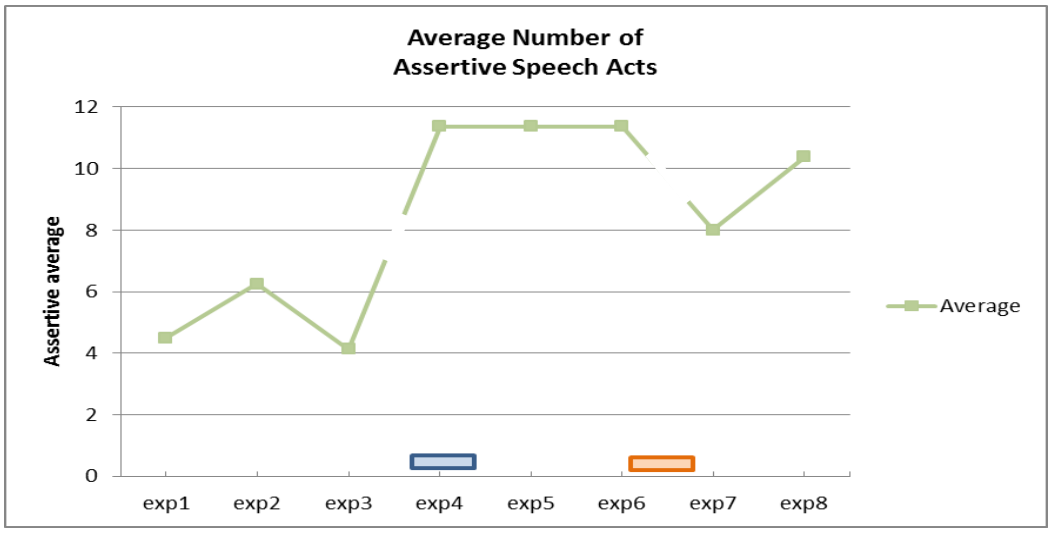


Figure 58 Assertive Average across the 8 experiments

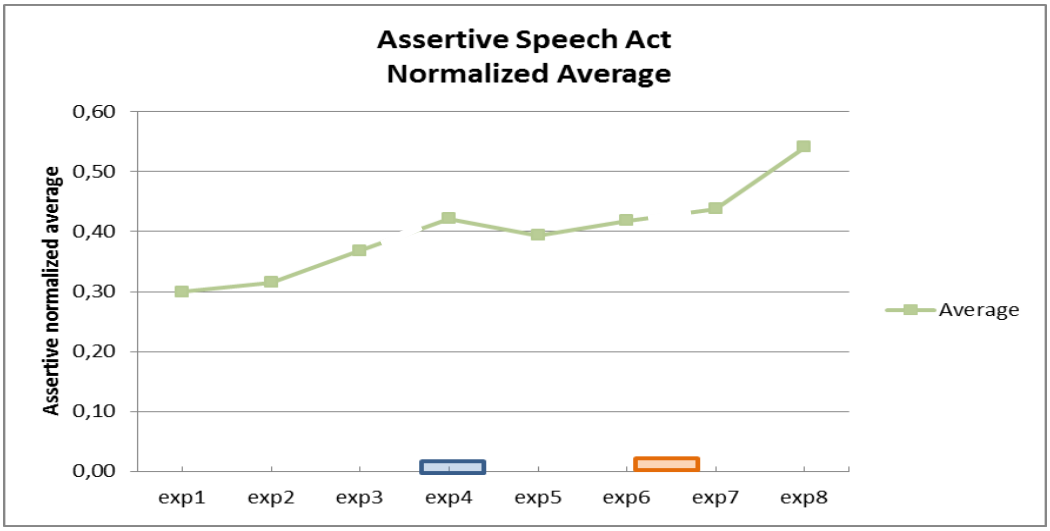


Figure 59 Assertive Normalized Average

The increase in the Assertive speech act category is verified as **highly significant** by a Repeated Measures ANOVA for both the token count and the normalized data. The change of the **total number of Assertive speech acts** across the eight experimental sessions was **significant** under the Greenhouse-Geisser correction ($F(2.4, 16.7)=3.51$, $p=.047$, $\eta_p^2=0.334$). This significant result is due to a significant linear trend ($F(1, 7)=13.79$, $p=.008$, $\eta_p^2=0.66$).

According to Friedman's Test, there was a statistically significant change in the Assertive speech act category across the eight experiments ($\chi^2(7) = 25.140$, $p = .001$).

The change of the **normalized number (use ratio) of Assertive speech acts** across the eight experimental sessions was **significant** ($F(7, 49)=3.50$, $p=.004$, $\eta_p^2=0.333$). This significant result is due to a significant linear trend ($F(1, 7)=19.59$, $p=.003$, $\eta_p^2=0.74$).

According to Friedman's Test, there was a statistically significant change in the normalized number (use ratio) of Assertive speech act category across the eight experiments ($\chi^2(7) = 19.242$, $p = .007$).

This increase will be discussed as the main indicator of saturation of the communication system and the increase in the skills of participants in using it. Assertive speech acts are the prototypical speech act that is used to represent and transfer objective information about the state of things, events or people.

4.4.2. DIRECTIVE SPEECH ACT CATEGORY

This category is a good example to show the help of normalization to visualize the variance across couples (communication systems) and the variance over time. The token count of all 8 couples' data for the Directive speech act category is visualized in Figure 60 below:

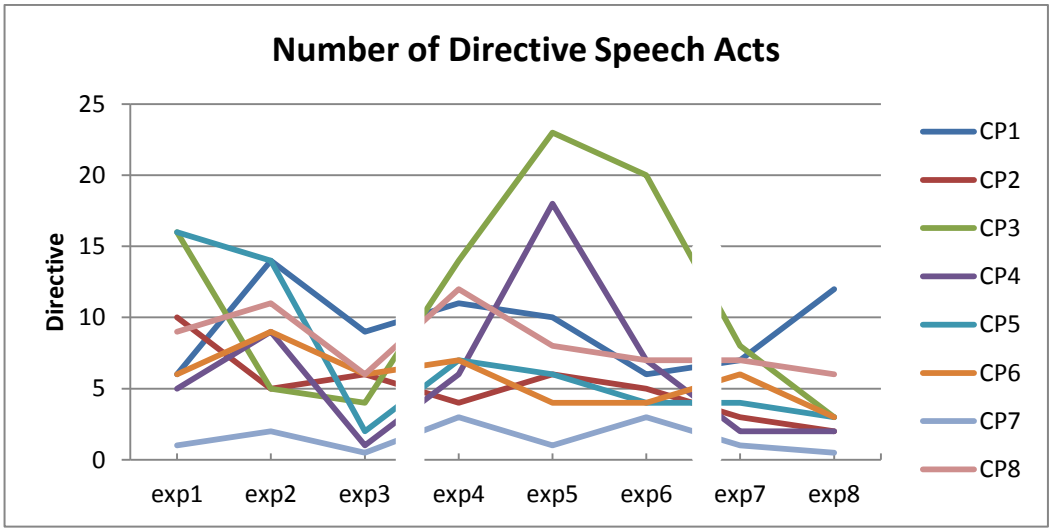


Figure 60 Number of Directive speech acts for the 8 couples across the 8 experiment

When each of *the token number of Directive speech acts,* for a couple in an experimental session, is divided by the total number of turns (speech acts), the effect of total amount of speech is removed. This is the normalization procedure used in this study:

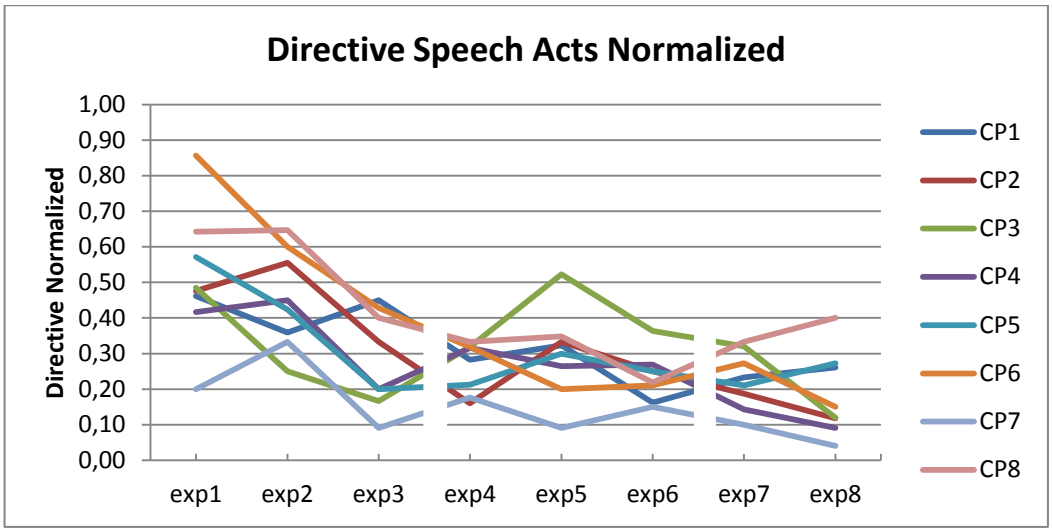


Figure 61 Directive Speech Acts Normalized for the 8 couples across the 8 experiments

This category of speech acts was counted first as tokens and was then normalized. The trends for the averages are depicted in Figure 62 and Figure 63, respectively:

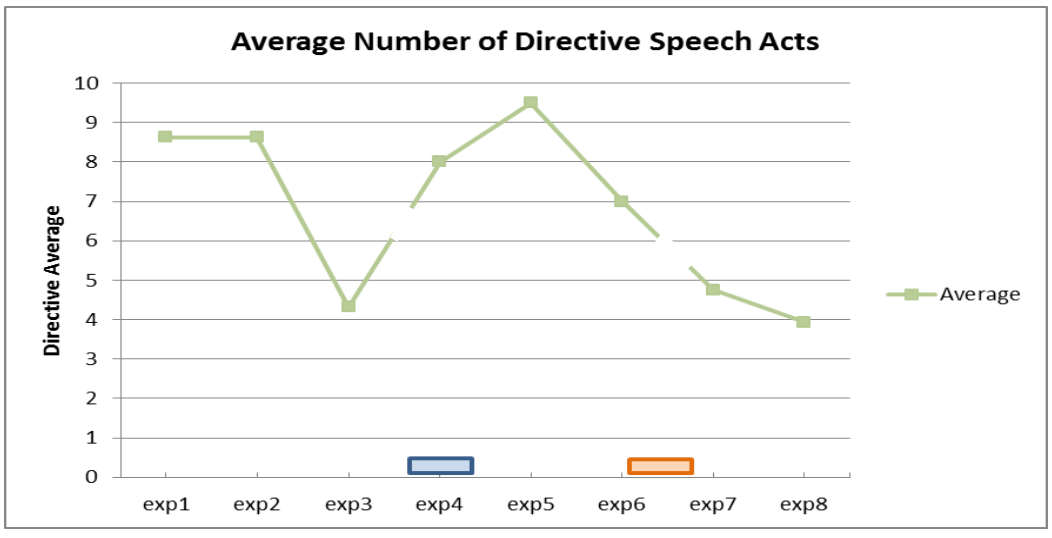


Figure 62 Average Number of Directive Speech Act Category across the 8 experiments

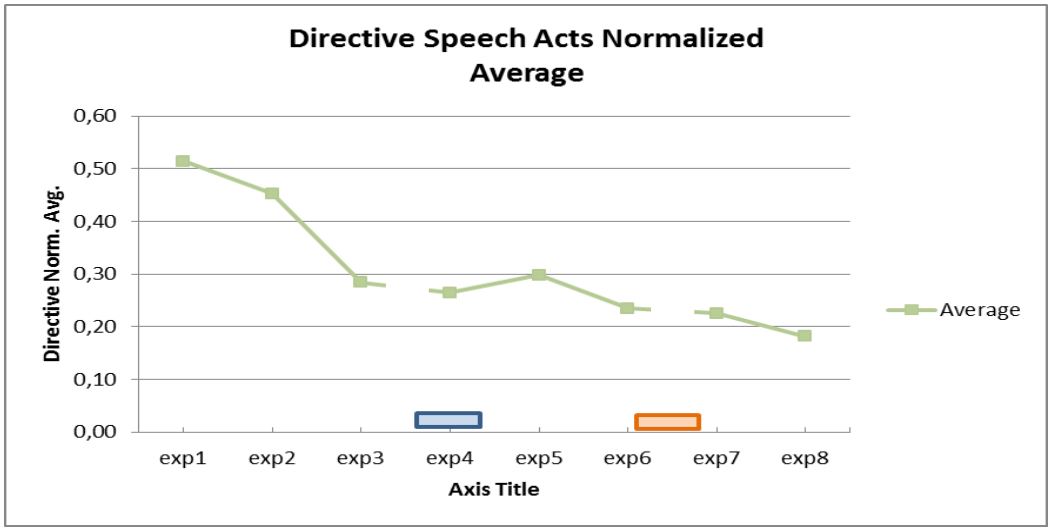


Figure 63 Average of Directive Speech Acts Normalized across the 8 experiments

There is a visible decrease in the token number of directives and normalized ratio of uses over total turn numbers in each experiment in the later sessions. The Directive speech acts are observed more when there is more need to coordinate actions for navigation and information transfer. This is more often the case in earlier sessions; however, when the skill and the communication system develop they seem to be needed less.

The Repeated Measures ANOVA for the Directives in terms of **token count** shows a **significant decrease** whereas the same analysis with the **normalized average data shows a highly significant decrease** over the eight experimental sessions.

The change of **the total number of Directive Speech Acts** across the eight experimental sessions was **significant** under the Greenhouse-Geisser correction ($F(2.3, 16.2)=4.95$, $p=.018$, $\eta_p^2=0.414$). This significant result is due to a significant linear trend ($F(1, 7)=15.87$, $p=.005$, $\eta_p^2=0.69$) and a significant quartic (fourth order) polynomial trend ($F(1, 7)=13.20$, $p=.008$, $\eta_p^2=0.65$)

According to Friedman's Test, there was a statistically significant change in the Directive speech act category across the eight experiments ($\chi^2(7) = 31.414$, $p < .0001$).

The change of the **normalized number (use ratio) of Directive Speech Acts** across the eight experimental sessions was **significant** ($F(7, 49)=10.92$, $p<.001$, $\eta_p^2=0.609$). This significant result is due to a significant linear trend ($F(1, 7)=31.32$, $p=.001$, $\eta_p^2=0.817$).

According to Friedman's Test, there was a statistically significant change in the normalized Directive speech act category across the eight experiments ($\chi^2(7) = 31.414$, $p < .0001$)

4.4.3. EXPRESSIVE SPEECH ACTS

This category of speech act was counted first as tokens and was then normalized. The trends for the average number of Expressive speech acts and the normalized average numbers are depicted in Figure 64 and 65.

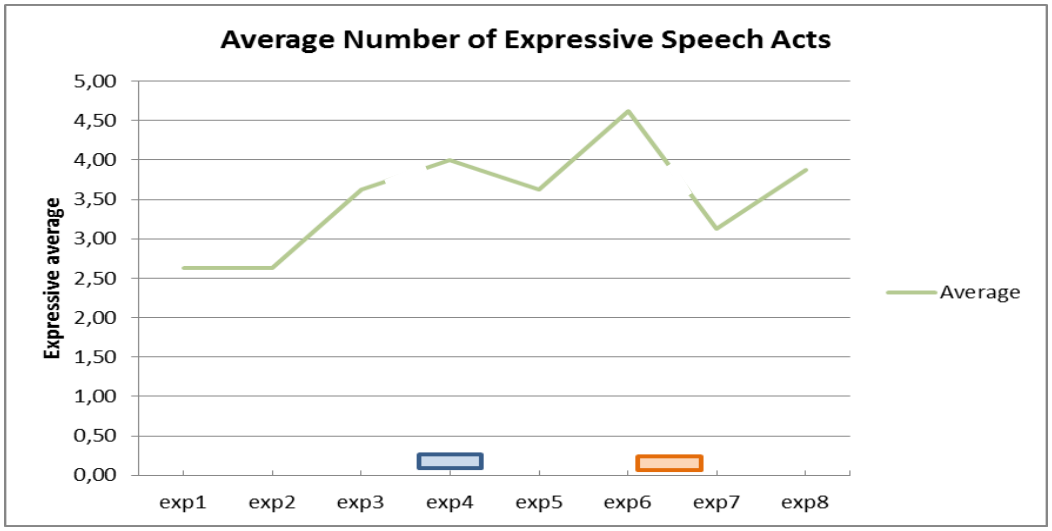


Figure 64 Average number of Expressive Speech Acts across the 8 experiments

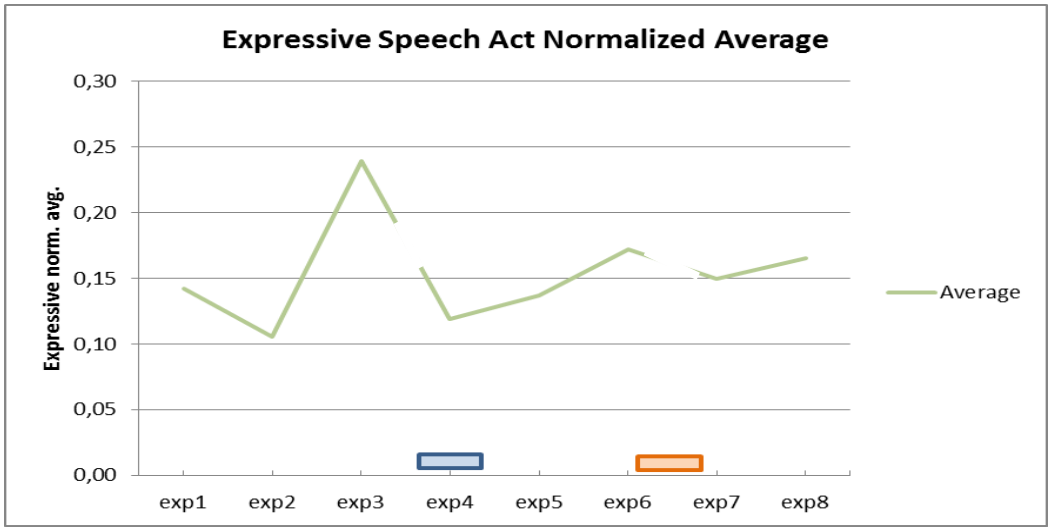


Figure 65 Average number of Normalized Expressive Speech Acts across the 8 experiments

The Repeated Measures ANOVAs for both of the data versions do not indicate any significant change during the eight experiments.

A Repeated Measures ANOVA revealed that the total number of Expressive Speech acts did **not** change **significantly** over the course of the eight experiments ($F(7,49)=0.83$, *p=.572*, $\eta_p^2=0.105$).

According To Friedman's Test, the change in the number of Expressive speech acts across the eight experiments **was not significant** ($\chi^2(7) = 8.654$, $p= .278$).

A Repeated Measures ANOVA revealed that the normalized number (use ratio) of Expressive Speech acts did **not** change **significantly** over the course of the eight experiments ($F(7,49)=1.26$, *p=.292*, $\eta_p^2=0.152$).

According to Friedman's Test, the change in the normalized number of Expressive speech acts across the eight experiments **was not significant** ($\chi^2(7) = 11.023$, $p= .138$).

This result will be discussed later but at this moment we can conclude that participants' relative use of Expressive speech acts does not change in comparison with the remaining categories.

4.4.4. COMMISSIVE SPEECH ACTS

This category of speech acts was counted first as tokens and was then normalized. The trends are depicted in

Figure 66 and Figure 67 below:

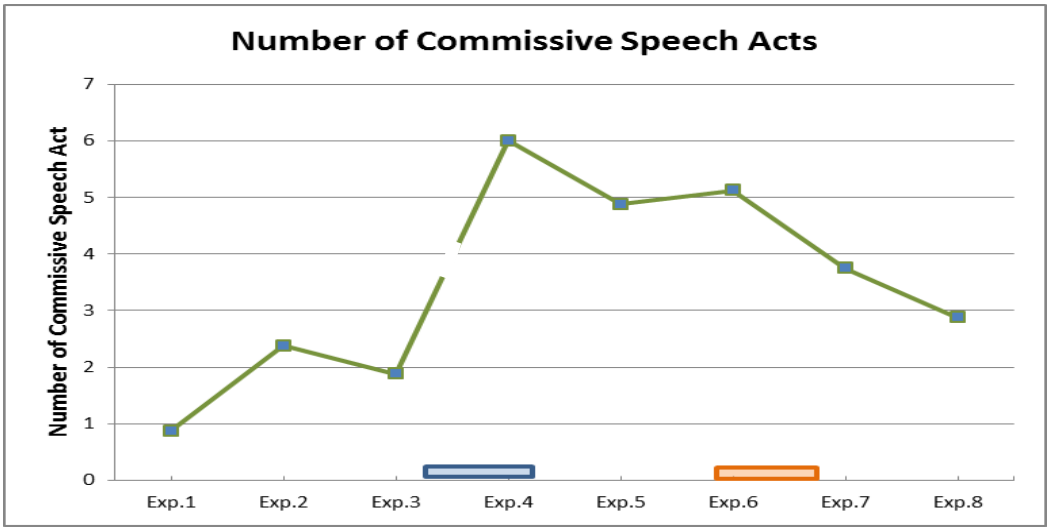


Figure 66 Number of Commissive Speech Acts

The Commissive speech act category shows an increase and then a more visible decrease after the 4th experiment where the 2 target task structure became more settled. This development can be argued to be due to the effective use of Assertive speech acts to transfer information about targets and couples' situations and less need for communication for navigational coordination.

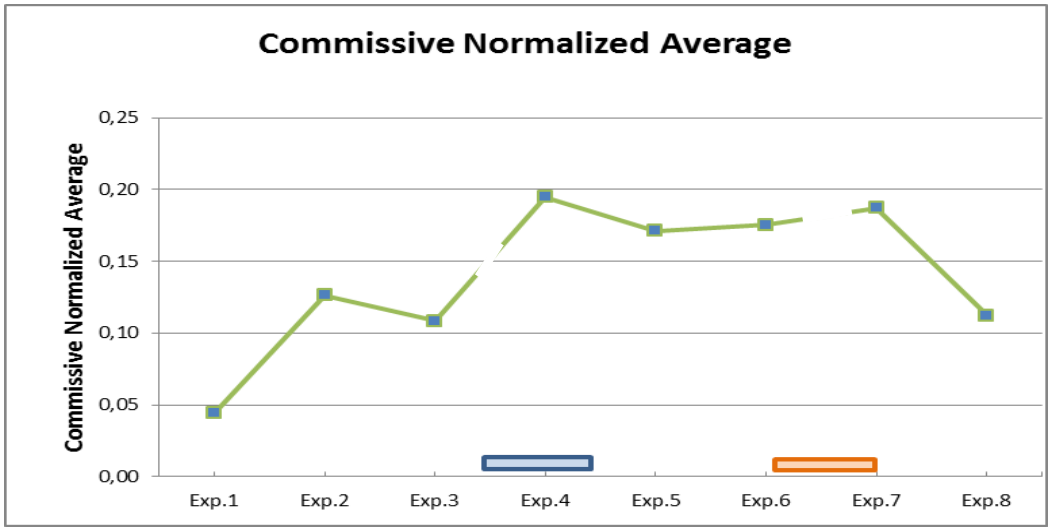


Figure 67 Commissive Normalized Speech Acts

A Repeated Measures ANOVA revealed that the change of **the total number of Commissive Speech Acts** across the eight experimental sessions was highly **significant** ($F(7, 49)=5.51$, $p<.001$, $\eta_p^2=0.440$).

According to Friedman's Test, there was a statistically significant change in the Commissive speech act category across the eight experiments ($\chi^2(7) = 27.399$, $p < .0001$)

A Repeated Measures ANOVA revealed that the change of **the normalized number (use ratio) of Commissive Speech Acts** across the eight experimental sessions was significant under Greenhouse-Geisser correction ($F(2.97, 20.8) = 4.05$, $p=.021$, $\eta_p^2=0.366$).

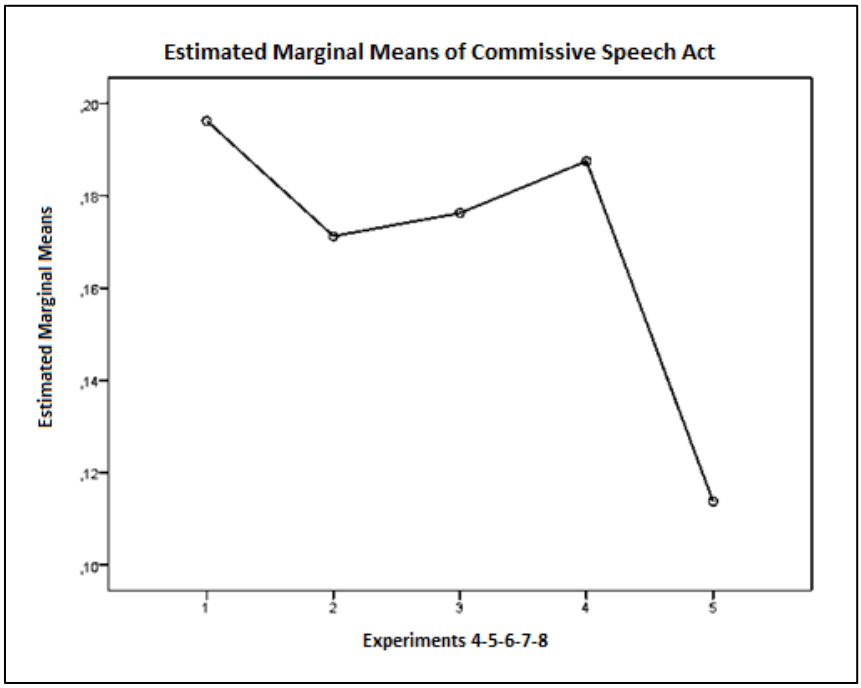


Figure 68 The Average Normalized Number (use ratio) of Commissive Speech Acts (in Exp4-5-6-7-8)

The decreasing trend during the process of saturation of communication system after the 3rd experiment (4-5-6-7-8th experiments) was tested by Repeated Measures ANOVA for the last 5 time points.

The change of **the normalized number (use ratio) of Commissive Speech Acts** across **the last five experimental sessions** was **not significant** ($F(4, 28)=1.535$, $p=.219$, $\eta_p^2=0.180$).

4.4.5. THE OVERALL ANALYSIS OF SPEECH ACT CATEGORIES

The combined speech act categories data, in terms of averages of token counts of categories and normalized data revealed differences between the various speech act categories but similarities between the token counts and the normalized data, see Figure 69 and Figure 70 below:

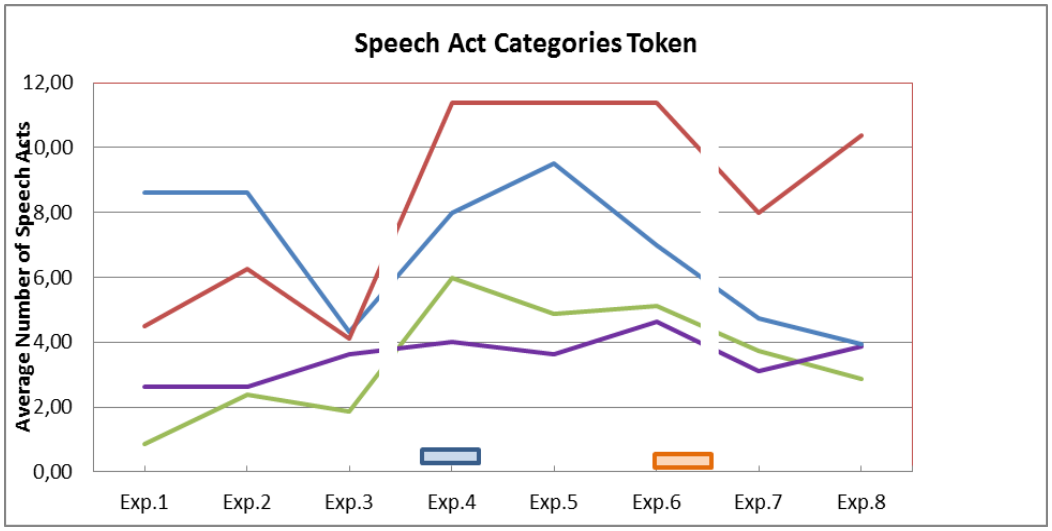


Figure 69 Speech Act Categories in average total numbers of 8 couples

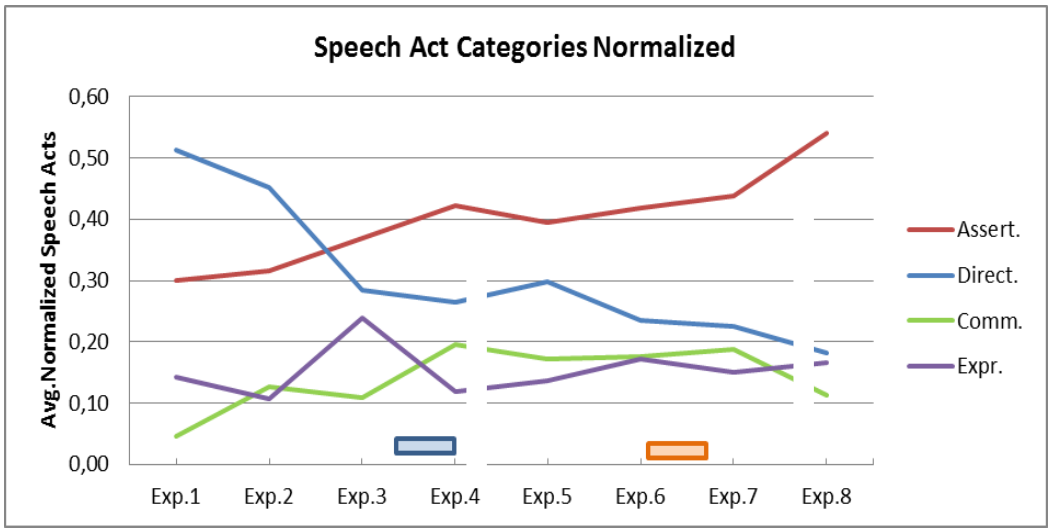


Figure 70 Speech Act Categories in average (use ratio) of 8 couples

The normalized data graph indicates opposite trends for Assertive and Directive speech acts after the point (exp4), where the new task structures had been settled. After this point, under a constant exposition to slightly different tasks and slightly different spatial conditions, couples developed their skills and collaboration on building a communication system.

The high use ratios of Assertives and relatively low use ratios of Commissive speech acts are plausible criteria to predict phases of increased efficiency of a communication system, where its lexicon and pragmatic power are sufficient for meeting the expected communicative challenges and succeeding on the collaborative tasks.

4.5. GENERAL DISCUSSION ON THE 4 LEVELS OF PARAMETERS

General Results for the Quantitatives:

The signature of the typical development of the eight communication systems is presented below:

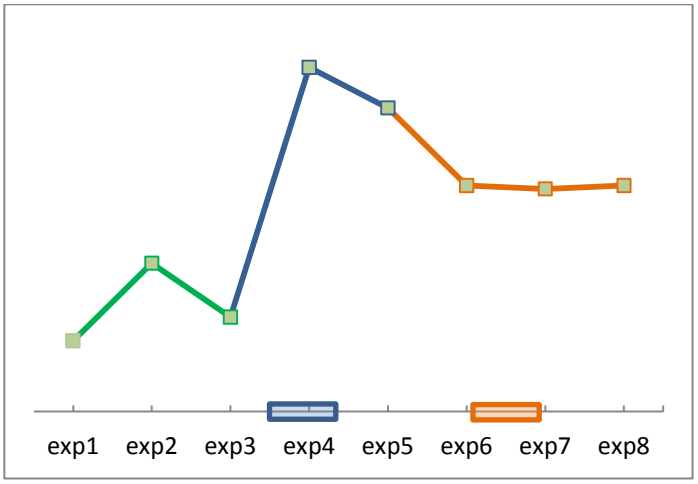


Figure 71 Generalized graph depicting the typical development of communication systems across the 8 experiments

The multi-color series line in the above abstract graph characterizes many common aspects of the development of these communication systems (in terms of their quantitative parameters). The parameters may be the total number of lexical items, turn numbers, token number of nouns or verbs or token number of turns that can be categorized as assertive speech acts. The graph displays the typical pattern of an initial increase followed by a decrease (phase 1), then a jump followed by a decrease again (phase 2) and then two steady last sessions (phase 3).

The green line segments indicate the initial phase of adaptation to the experimental constraints (phase 1). This happens usually in the second and the third sessions. So communication and task performance are easy in the third session, the tasks are achieved by means of a relatively low amount of communicative and physical action (physical in the virtual world, of course).

Then the intervention comes, as designated by the blue horizontal band on the time level of exp4 on the horizontal axis (phase 2). The increase in the number of the task targets from one to two creates a disturbance in the system and a process of re-adaptation is realized which is represented by the respective blue line segments for the 4^{th} and 5^{th} experimental sessions.

The orange vertical band between exp6 and exp7 time levels designates the two months break between the experiments 6 and 7. Despite its length, the break appears to be not influential in causing a decay in the participants' performance as we can track by means of examining the trends of change in the several parameters during the exp6-exp7-exp8 period.

The resemblance in the adaptation and re-adaptation patterns between the green exp1-exp2-exp3 line segments and the blue exp3-exp4-exp5 line segments is apparent. The scale difference (i.e., the higher values in the second phase) is the effect of the amount of communication in the re-adaptation phase.

At this point another set of characteristics of the communication system can be demonstrated by removing the effect of amount of communication and looking at the use, success and participation ratios or type counts of elements of the communication system.

For example, the averages of the total number of certain speech acts in each session are depicted in Figure 69 in the previous section. These numbers increases 2 or 3 times more in the 4^{th} experiment; however, when we filter out the effect of the amount of communication and follow the trends of the relative use ratio of speech acts we see that the changes of these trends are smooth and steady

from the 3rd to the 4th and 5th experiment (see Figure 70). We can claim that there are underlying processes of change for certain parameters which are only prone to be affected by the condition of repeated exposition to qualitatively similar environments and communication requirements. The underlying process is the increase in the level of application of existing communicative skills in the experimental context, by means of interaction among the dyad members.

This holds true for the level of lexical categories as well: in the two Figure 54 and Figure 56, (reproduced here for convenience) ,the first is for the token count (which is affected by the amount of communication) and the second is the normalized one:

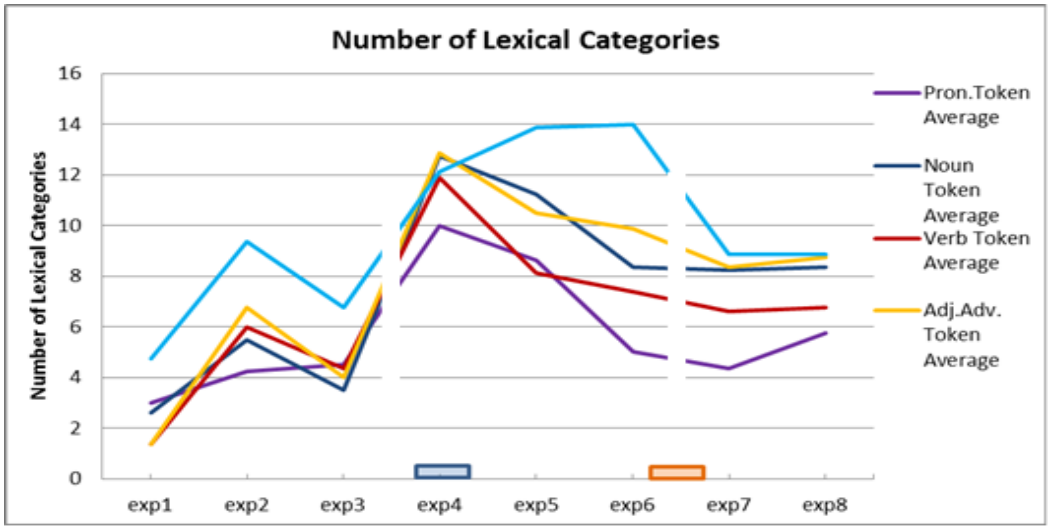


Figure 54 reproduced here for convenience (The average token number of lexical categories across 8 experiments)

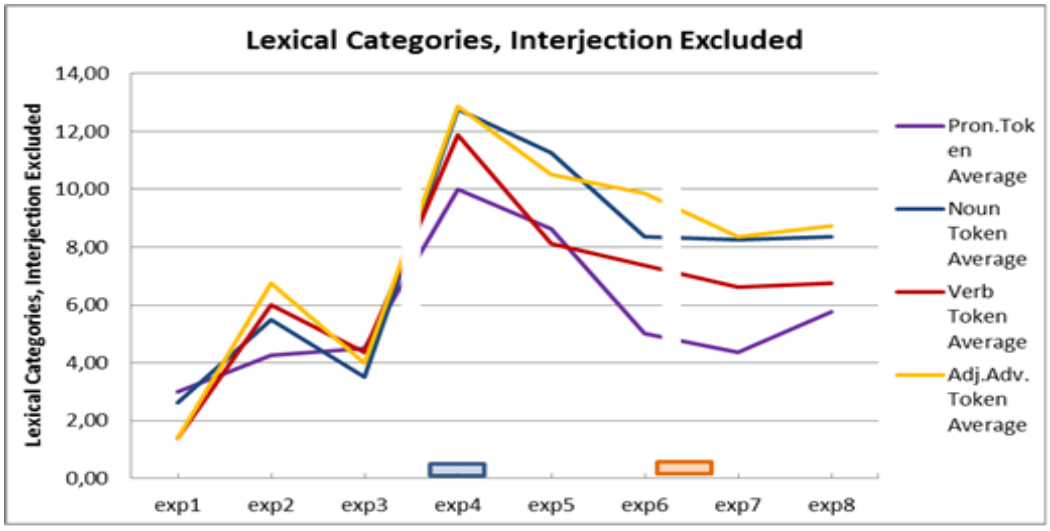


Figure 55 reproduced here for convenience (The average normalized number (use ratio) of lexical categories across 8 experiments)

For each category the use ratio of the lexical category is changing but it is not changing in the same way as it does for token count graph (Figure 24) between exp3 and exp4.

We do not see this resilience of trends against the change in the 4th experiment for the purely quantitative parameters analyzed in the first section, because they are about quantities by definition. They cannot be normalized like lexicals and speech act categories.

The only exception in the quantitative parameters section is the Turn Success ratio (Figure 22). This parameters' being a ratio makes is possible to test the prediction that these underlying processes of increasing skill or experience of participants and the efficiency of the system are unaffected by the first intervention.

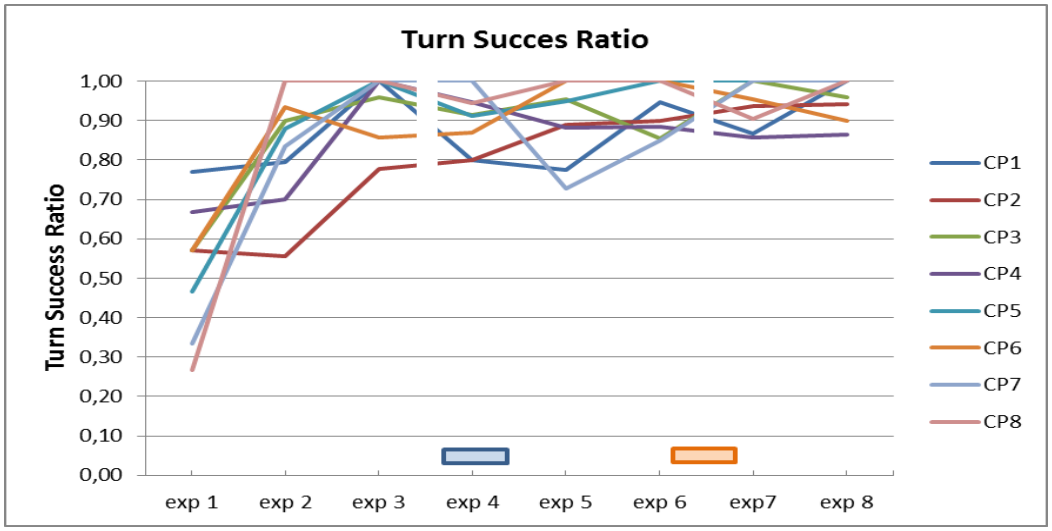


Figure 22 **Turn Success Ratio,** reproduced here for convenience

The same discussion applies to the Complexity score section in its dedicated overall analysis.

Similarly, universal smooth trends or abrupt interruptions may be examined for the change between the 6th and the 7th and 8th experiments (phase 3). There is a break around 2 months between exp6 and exp7. This interval may cause a deterioration in the skill of using the emerged communication system, i.e., using the established lexical inventory, general strategies, etc.

The same graphs and statistical analysis shows, however, no (or only slight) changes on the proposed parameters underlying the development and the efficiency of the communication system. The trends for *the New Lexical Items, Number of Turns, Turn Success Ratio, Type over Token Lexical Items Ratio, Normalized Complexity Score, Normalized Assertive Speech Acts* were not (or hardly) affected by the two months gap between the 6th and the 7th experiments.

There is only a change (an increase) for the average time for the 7th experiment where the familiarity with the artificial environment is needed to be regained. The subsequent decrease of time scores for the 8th session may be an indication of this fact.

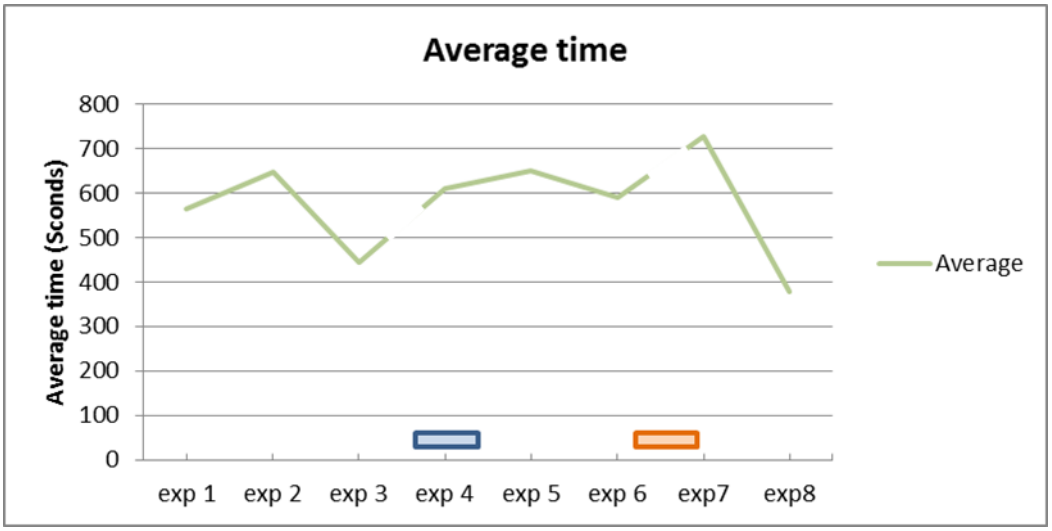


Figure 19 Average Time, reproduced here for convenience

4.6. RESULTS OF THE QUALITATIVE ANALYSIS OF THE EXPERIMENTAL DATA

In this section the qualitative features or aspects of the emergence of a new communication system are examined. The main question addressed is "**How did the participants, the dyads, figure out to communicate under the given constraints?**"

Before attempting to give the answer, it is necessary to state that nearly all of the participants expressed that they doubted their own capacity for achieving the general task of the experiments as a communication game. The experiment, the rules and the requirement were presented to them by a simple 1 page document and also explained personally by the experimenter. This reaction after the instruction may also be a kind way of expressing their doubt about the experimental design. After all, these people were volunteers and were investing their and the other participant's 8 hours of lifetime dispersed to at least 4 different days for 8 experimental sessions. They may also have been afraid being stuck in an experiment for boring hours. So the initial impressions of the participants were that it was not an unchallenging mission.

Interestingly, and contrary to their own initial doubts, all of the couples succeeded to communicate effectively to a certain degree at the latest in the third experimental session. Their motivation increased and they "enjoyed the game". At least one of the members of the dyad figured out the main strategy to exploit their shared and pre-existing communicative skills and NL competence.

4.6.1. THE COMPRESSION OF NL UTTERANCES INTO A GIVEN COMMUNICATION TOOL

The main strategy was **compressing the NL words into the existing constrained set of letters** and symbols. The available letters q, w, e, r, t, y, u, ı, o, p and symbols →, "?" provided them with a means to coin lots of lexical items which show similarities with the words required for the context-relevant utterances in the initial phases of the experiments. For example, the "??" worked as a general interjection and expressive speech act to question the situation ("what are we doing here?"), to check the attention or availability of the partner, or to express frustration if one could not understand the interlocution. In the first two sessions the most principal pronouns, wh-words "wer?" and task-related nouns and adjectives were tried to be composed by means of these available building blocks. Some of the missing letters necessary for the required words were substituted by available similar ones. Similarity was exploited both in terms of scripture and phonetic similarity. Majuscule i was used for miniscule L "qeI" or "qeı" (visual similarity). In the second case to use of the word "küre", "qre" was coined and comprehended successfully in most of the cases. Here phonetic similarity between q and k was used to extend the scope of available word-building items (letters).

The third use of available letters capitalized on their iconic properties, for example, the letter I is used to create numbers "ı, ıı"; the letter "O" is used to denote the target object sphere, "*y O*" ("green sphere"); → the arrow sign used as a general orientating, directional, relational lexical item: "*ı → yO*" ("I am going to

green sphere"), "*our trqt -> qup*" ("*our target is cube*"), "*prowr -> quereqı*"(brown means "kahverengi").

Table 19 Sample Communication Transcription of Couple 8

O. D. CP8												
	EXP1				EXP 2				EXP 3			
	turn #	Time	Partcpt	Message	turn #	Time	Partcpt	Message	turn #	Time	Participant	Message
	1	03:56:06	partcp2	q	1	04:37:11	partcp2	??	1	05:13:27	partcp2	were re you
	2	03:56:12	partcp1	?	2	04:37:39	partcp1	wer r you	2	05:14:05	partcp1	ret pyrt
	3	03:56:16	partcp2	q?	3	04:38:02	partcp2	pyr	3	05:14:16	partcp2	weıt
	4	03:57:17	partcp1	w	4	04:38:23	partcp2	purpıe pyr	4	05:14:56	partcp1	weıt you
	5	03:58:49	partcp2	?	5	04:39:16	partcp1	wer we qo	5	05:15:11	partcp2	weıt weıt
	6	04:01:17	partcp1	w	6	04:39:23	partcp2	wer?	6	05:15:38	partcp2	you->ı
	7	04:01:30	partcp1	w	7	04:40:00	partcp2	ı->you	7	05:15:41	partcp1	ı->you
	8	04:01:44	partcp2	?	8	04:40:09	partcp1	yee	8	05:15:50	partcp2	qo qo
	9	04:02:54	partcp1	"->w	9	04:40:43	partcp1	wer r you	9	05:15:51	partcp2	ye
	10	04:03:06	partcp1	w	10	04:40:54	partcp2	purp pyr	10	05:16:49	partcp2	oryt
	11	04:05:15	partcp2	wer r you?	11	04:41:04	partcp1	weıt	11	05:17:00	partcp2	qree pyr
	12	04:05:18	partcp1	p	12	04:41:13	partcp2	oryt	12	05:17:05	partcp1	trqet?
	13	04:06:41	partcp2	?	13	04:41:52	partcp1	->	13	05:17:10	partcp2	ye
	14	04:06:54	partcp1	w->	14	04:42:18	partcp2	qo	14	05:17:20	partcp1	ooo yeeee
	15	04:07:17	partcp1	w	15	04:44:46	partcp1	->	15	05:17:24	partcp2	yeee
					16	04:46:34	partcp1	->				
					17	04:54:58	partcp2	were to?				

As it can be seen in Table 19, in the sample transcription of Couple 8, after an initial trial for the use of single characters to denote words, the participants noticed that NL-like words can be created and it became the main solution for the communication.

Iconic and deictic use of → is also another strategy. Most of the dyads fell back on this trivial strategy (CP1, CP2, CP3, CP5, CP6, CP8).

4.6.2. THE PARSIMONIOUS SINGLE CHARACTER STRATEGY:

One of the dyads, Couple 7 (CP7) managed to use the tactic of *using single characters as words*, successfully across the eight experiments. This was possible by means of very scarce communication, lots of visual coordination, effective task-sharing and collaborative navigation, and heavily relying on the inference of

meaning based on the relevance of very parsimoniously used lexical items to the context. Another issue that made this possible is that the participants' expectation space narrows due to experience and familiarity. The participants see the environment and notice that there are finite types of objects and colors and color regions. This narrows down the possible intentions and meaning behind a very parsimonious utterance. Similarly, as the tasks are nearly uniform across the 8 experiments, they can develop expectations about the nature of the tasks that may be given to the other participant: The next target will include an object with certain known shape and color. So, when the uniformity of the task structure is confirmed, not only the objects and the features, but the possible speech acts are also very limited to process the relevancy of the utterance. For example, "pp" in Exp6 (in Table 20 below) is very unlikely to be about two pyramids but rather a purple pyramid.

In the following sample transcript from the last the 6th, 7th and 8th experiments, we can see that the strategy of adopting single characters for words is successfully implemented. Interestingly, the rare attempt of using the word "tree" (a possible one with the given characters to express "three", "tree" and sometimes "there") had failed since the partner did not comprehend the meaning and the intention of the speaker.

In this couple, "y" is both used to denote "yes" and "yeşil"; since "O" (iconic) is used for sphere, it can be inferred that in this context "q" does not denote *qure* (küre, sphere) but "cube" ("küp"). All these instances of ambiguity arising from polysemy or homonymy are resolved by virtue of the optimal relevance of the desired intention assumed by the hearer (i.e., according to the cognitive principle of relevance (Wilson, 1999)).

Table 20 Sample Communication Transcription of Couple 8, the parsimonious single character strategy

Couple 7 CP7_T_O														
Experiment 6					Experiment 7					Experiment 8				
turn #	Time	Participant	Message	S/F	turn #	Time	Participant	Message	S/F	turn #	Time	Participant	Message	S/F
1	12:30:52	partcp2	tree	Fail	1	12:38:30	partcp2	yq	Success	1	01:01:53	partcp2	yp	Success
2	12:30:58	partcp1	y	Success	2	12:39:02	partcp1	po	Success	2	01:02:13	partcp1	yp	Success
3	12:32:31	partcp2	r tree	Fail	3	12:39:36	partcp2	yq ->	Success	3	01:03:21	partcp2	yp	Success
4	12:34:23	partcp2	- p p	Success	4	12:41:34	partcp1	yq->	Success	4	01:03:47	partcp2	yp->	Success
5	12:34:26	partcp2	p p	Success	5	12:42:02	partcp2	y	Success	5	01:04:05	partcp1	yp->	Success
6	12:36:33	partcp1	p p->?	Success	6	12:42:11	partcp1	y	Success	6	01:04:07	partcp1	y	Success
7	12:36:50	partcp2	y	Success	7	12:42:20	partcp1	po	Success	7	01:04:13	partcp2	y	Success
8	12:36:56	partcp1	y	Success	8	12:46:14	partcp2	po->	Success	8	01:04:19	partcp1	yo	Success
9	12:37:04	partcp1	e q	Success	9	12:46:39	partcp1	po->	Success	9	01:04:25	partcp2	yo	Success
10	12:37:15	partcp2	p p ?	Fail	10	12:46:42	partcp1	y	Success	10	01:05:38	partcp2	yo->	Success
11	12:38:12	partcp2	p p ->	Success						11	01:05:38	partcp1	yo->	Success
12	12:38:36	partcp1	y	Success						12	01:05:42	partcp1	y	Success
13	12:38:38	partcp2	y	Success										
14	12:38:41	partcp1	e q	Success										
15	12:38:43	partcp2	e q	Success										
16	12:39:50	partcp1	e q - p q	Success										
17	12:43:57	partcp2	e q ->	Success										
18	12:44:03	partcp1	p q->	Success										
19	12:44:10	partcp2	y	Success										
20	12:44:16	partcp1	y	Success										

The significance of this strategy is that even though it includes shortcuts to lexical items of Turkish and English, we cannot identify/distinguish any additional syntactic feature of Turkish and English except the head-last noun clause structure. It can be claimed that this communication system just represents the discourse structure of the dyad's communication by only borrowing the common NP syntactic structure of the two common languages that the participants share.

Another issue is that, even though we have a radically different, parsimonious strategy, distinguished from other strategies of compressing NL into the experimental constraints, in general, the (normalized) quantitative trends and features are similar to the communication systems created by other couples, when one looks at the trends of type over token ratio, turn taking ratio, turn success ratio, etc.

4.6.3. SUBSTITUTING THE REDUNDANCY

A third dyad, Couple 4 used another strategy to denote required lexical items by means of given characters. This strategy is characterized by utilization of the symbol ? to substitute missing characters. In this way, they did not need to

bother themselves, in most of the cases, to discover graphic or phonetic resemblances between the available communication tools and the required ones.

Table 21 Sample Communication Transcription of Couple 7

Experiment 7					Experiment 8				
turn #	Time	Participant	Message	S/F	turn #	Time	Participant	Message	S/F
1	05:55:26	partcp1	e???tu? qure	Success	1	06:37:07	partcp1	ye?ı? qure	Success
2	05:55:57	partcp2	ye?ı? qup	Success	2	06:37:14	partcp2	ye?ı? pyr??ıt	Success
3	05:56:20	partcp1	oq	Success	3	06:37:19	partcp2	oq	Success
4	05:58:30	partcp1	w?ıt	Success	4	06:37:20	partcp1	oq	Success
5	05:59:09	partcp2	yee,	Success	5	06:38:22	partcp1	ye?ı? pyr??ıt	Success
6	05:59:10	partcp1	oq	Success	6	06:38:26	partcp1	?ere	Fail
7	05:59:12	partcp1	yee	Success	7	06:38:44	partcp2	?	Success
8	05:59:22	partcp2	your	Fail	8	06:38:44	partcp1	q?were??ı	Success
9	05:59:25	partcp2	?	Fail	9	06:38:57	partcp2	yoq	Success
10	05:59:42	partcp1	e???tu? qure	Success	10	06:39:08	partcp2	ye?ı?	Success
11	05:59:52	partcp2	qo	Success	11	06:39:22	partcp1	yee	Fail
12	06:00:00	partcp1	oq	Success	12	06:39:29	partcp2	weıt	Success
13	06:01:25	partcp1	yee	Success	13	06:39:32	partcp1	ıt? ?ere	Success
14	06:01:30	partcp2	yee	Success	14	06:39:33	partcp1	oq	Fail
					15	06:40:47	partcp2	oq	Success

As it can be seen in the Turns 1, 2, 4, 10 of exp7 and 1, 2, 5, 8, 10 and 13 of exp8 the nouns and adjectives in syntactically more complex utterances are expressed by substituting the missing characters of the necessary lexical items. Only simple interjections and pre-established pronouns, in the single lexical items utterances are exceptions to this general strategy. This strategy "cheats" the constraints of the experimental design on the communication channel by exploiting the redundancy of characters of the NL lexical items in the written form. Again similar with CP7's strategy of single character words, *contextual* and *relevance-based* inferences help with the identification of intended lexical item in NL. For example, by reading a lexical item like "ye?ı?" or "e???tu?", since the context, environment and task structure requires some color adjectives, the hearer can process these as "yeşil" (green) or "eflatun" (purple). This case supports the significance of processing the relevance of the utterance for correct comprehension.

Still the excessive use of the symbol “?” makes it very ambiguous in the context of interrogative uses. “ye?ı??” may be intended as a declarative or express the question “Is it yeşil?” in which case it is very hard to identify as “yeşil??”. The excessive use of “?” obscures the obvious interpretation as a question mark when it is used for its conventional function.

4.6.4. THE SEARCH IN TWO NL LEXICONS TO FIND COMPRESSIBLE NL LEXICAL ITEMS: EXPLOITING THE SECOND LANGUAGE:

A fourth common feature of all the previously mentioned strategies is the exploitation of shared NL competences. Eight of nine couples are bilinguals who learnt English as second language. These couples heavily relied on English lexical items, expressions and syntactic structures when the available characters and context provided an advantage that way. A more universal trait is that nearly all the complex clauses and utterances are a blend of syntax and lexicon of Turkish and English, as in the example below:

(“yeqı qp wr?”, (“yeşil küp nerede?”))

Interestingly, the members of those Couple who had medium or elementary level competence in English also used English words and sometimes syntactic structures. This may be due to the selection of available characters, namely “**qwertyuıp**”, which contains q and w as non-Turkish characters, (they are also used to substitute k and v). Furthermore, the shared English-based, computer-mediated communication, Internet culture conventions, (“oq” as a universal), the institution METU, and the general computer/gaming environment may also have primed the use of English.

As a limitation of the study there had been no chance of finding additional couples and trying different (maybe less) numbers of different sets of characters (like **asdfghjkl**) and compare the outcome with the present study.

4.6.5. THE UTILIZATION OF WRITTEN COMMUNICATION AND INSTANT MESSAGING ENVIRONMENT:

a-) The persistence of utterances and consecutive/ adjacent turns during a conversation makes it easy for the hearer to process real-time communication and make inferences, re-evaluate the possible meaning of an utterance or lexical item of a past conversation in that experimental session.

b-) On the speaker side, the linear and persistent nature of written messages provides the possibility of using the linear/spatial organization and the graphic existence of lexical items and symbols

("you –> me", "you follow me"); (since only the right arrow is available the leading person is written on the right and at the end of the sentence, if it was the left arrow available the preference might be in the form "Me ← you")

("qre → O?", ("qree and O means the same sphere?"))

("our tqrt qre → ?" ("Is this the sphere our target on the right?"))

c-) Exclamation like lexical items are created by repeated use of certain characters and "?"

"Wıot????" (The participants is asking or requesting attention and an instruction from the partner, since she is not given a target)

"Qreeeeee" ("This is the sphere we find it hurray!!" (the lexical item for sphere is qre))

d-) Consecutive turns are also used to indicate the order of tasks that will be attained collaboratively, like forming a persistent task list to remind the targets:

Turn 4 P1: Tqrt → prpı qre ("Target is purple sphere (*küre*)")

Turn 5 P2:Trq epıtu pyrt ("Target is purple (*eflatun)* pyramid")

4.6.6. GENERAL COGNITIVE SKILLS AND TRAITS REQUIRED AND USED TO ESTABLISH CONVENTIONS:

Even if there was an initial disbelief about the possibility of successfully communicating under the given constraints, starting from the very first

experimental session, the participants enjoyed the game and used their general communicative skills and experience to solve the puzzle of creating a new set of conventions (lexical items, a simplified syntax and discourse structure).

There may be some number of unsuccessful attempts of introducing new words or expressions, however, these are silently ignored, and communication and sometimes miscommunication flew smoothly. This expectation, the confidence in and the assumption of "I do not understand now, but it will be revealed later in this experiment" worked. In fact the experimental design helped, aimed and expected this. A few features and object categories are presented as targets and building blocks of the environment repeatedly, so that the participant managed to figure out the intention of the speaker in this or that context later. The experimental set-up thus provided to test the alignment of the use of introduced lexical items and their resilience in the lexical inventory. Without these repeated encounters the conventions would be single-use and disposable.

4.6.6.1. OPTIMAL RELEVANCE

The above skills and traits of the dyads can be related to two theoretical notions introduced previously in the second chapter, the literature review: one is "the communicative principle of relevance" and the second is "learning-by-using" introduced by Galantucci (2005).

The communicative principle of relevance is argued to be a working principle of an assumed *intentionality detector* or *intentionality tracker* sub-module of the mind-reading module. The principle sanctions [claims] an optimal relevance for every ostensive stimulus categorized as a communicative action by the receiver (Wilson and Sperber, 2006, p. 614). The receiver complies with this assumption of "the message's presumption of its own relevance" and does not rule out the relevance of the message with respect to the context of coordination, especially in the case of failed turns. This may be the reason of the persistency of a lexical item in

an unsuccessful turn, both on the speaker and the hearer side. In the data analysis phase of the present study, there is a stage of manually browsing and counting some selected items in the experimental data, using the logs of communication. In that phase, the condition for a new lexical item to be added to the lexical inventory of a dyad was that, when it was first used by the speaker, the hearer had to comprehend the message and the lexical item matching with the speaker meaning. Otherwise the lexical item is considered as an instance of unsuccessful attempt of introducing a new word to the system. However, even though the hearer may fail to detect the speaker's mapping or meaning to the signal and dismiss or ignore the message, it can be that that very hearer uses that unsuccessfully introduced lexical item in a second time (for example in next experimental session), successfully, and in the previous speaker's intended meaning. This means that there may be an ongoing processing of relevance attribution, even if the message is not comprehended and transformed into a conversation at its first appearance. The signal may be recorded in the cognitive system of the hearer as a part of a partially complete knowledge structure or frame. When the context of that signal is experienced and realized by the hearer, the failed comprehension may be completed in the next occurrence of the context and the signal and the lexical item may be circulated by the initial rejecter.

4.6.6.2. LEARNING BY USING

The second notion, *learning-by-using*, is the conscious or unconscious, trial-and-error-based attempt to increase the successful association of an un-comprehended signal with a context (Galantucci, 2005). When a signal cannot be assigned firmly to a relevant context or meaning, still some associations with some contexts are established, and after some successive encounters, the strength of this associations may increase or the association may be removed or a strict meaning of the signal can be established. This notion is not offered as a feature of the human

cognitive architecture as Sperber and Wilson do for the principles of their Theory of Relevance. Rather, learning by using is presented by Galantucci as one of many heuristics used when human agents try to convene upon some shared signals in a collaborative context. Crucially, learning by using is not a conscious application of trials and errors, or a kind of hypothesis testing; this learning is more like unconscious, experience-dependent mechanism based on the statistical relationships between the co-occurring contexts and signals, as in statistical learning (Xu and Garcia, 2008). Galantucci emphasizes the unawareness of the participants of the dynamics of their learning and refers to a study that investigates the unconscious experience-independent acquisition of infants, of word segmentation in fluent speech (Saffran, Aslin & Newport, 1996)

4.6.7. EXISTING COMMUNICATION SKILLS OF CONVERSATION

The participants not only exploited their existing access to the lexicon and the syntax of one or two NL's, but used their available skills of making conversation: their intuitive familiarity with the adjacency pairs, turn taking, solving turn taking problems, the overlaps in this synchronous mode of written communication in an instant messaging environment. Let us consider the repair mechanism in a case of misunderstanding and broken sequence of consecutive turns for one participant (P1):

Turn1 P1- tqrt we ("our target")
Turn2 P2- w r u? ("where are you?")
Turn3 P2- yqı qre ("yeşil(green) küre(sphere)")
Turn3 P2- trqt we yqı qre (" our target yeşil(green) küre(sphere)")

The initial turn of P1 got interrupted by P2 and the single utterance became distributed onto two turns (Turn1 and Turn3), thus losing its unity. As a self-repair mechanism (Schegloff, 1977) P1 repeats the utterance as a whole and P2

does not take the turn, is not confused and does not repeat his/her un-answered question (in Turn2).

4.6.8. IS THE NEW SYSTEM A COMPRESSED NL OR ARE THERE ANY UNIQUE FEATURES?

The case of Couple 7 shows that it may be the case, indeed, that they have created a set of communication conventions which was efficient for the adaptive requirements, and which saturated after a while and was thus a success story. These conventions are only initially introduced by NL, Turkish or English, but did not rely on its syntax and remained as abstract representations of target object declarations. In this sense, NL serves as a "scaffold" that provides some abstract framework to be filled in by idiosyncratic – but non-random – communication practices of each couple.

Secondly, some elements of the communication systems of the remaining couples are also distinctive. Not all of the elements of the emerging system may correspond to a NL lexical or syntactic category. They may have lexical items or even speech acts that may be considered as unique or may be composite in terms of meaning and/or functions. For example, textual instant messaging and limited linguistic building blocks leads the subjects to invent novel greeting, attention-grabbing and celebrating interjections like "oop", which is not exactly "hoop" in Turkish (used in the greeting sense by Couple 3). Similarly, the participants may mimic an NL interjection "yuppie" as a celebration and then use it in the form, "i yuppie ye qre" ("I have found/take care of green sphere"). , Here "yuppie" is a verb but it is not exactly the verb "find". It has its own connotations and history and can be used as something else in a different but relevant context, unlike the verb "find" can be used. Using polysemy as a strategy and avoiding ambiguity is about the same skill of communicative performance. Being accustomed to polysemy enables the

participants to comprehend closer but different connotations of a word and match the context and intention with it successfully.

CHAPTER 5

5. DISCUSSION AND CONCLUSION

5.1. THE COMMUNICATION SYSTEM

The development of the communication systems across eight experiments was a development indeed from a rudimentary initial level to a saturated, mature level, sufficient for the experimental task requirements. This development was observed by means of tracking the changing trends of several quantitative and linguistic/pragmatic parameters. These parameters and their observed trends may allow a comparative specification for the characterization of similar emerged/emerging communication systems which, in turn, may serve as a model for NL. The specification can be summarized from the results of the present study as follows:

1- A lexical inventory evolves in the system to meet the task requirements and the demands of the corresponding coordinative actions. The lexical inventory moves from an empty state to a relatively rich state swiftly, immediately after the speakers of the communication system have figured out and negotiated the production strategy of lexical items. The curve representing the trend of the counts of new lexical items added to the inventory in a time series is expected to have an inverse proportional, asymptotic shape, i.e., after an initial increase, the curve will converge on a certain number of lexical items sufficient for the complexity of the given task.

2- The system tends to support the coordination required for the successive tasks with similar difficulty, by production of smaller amounts of communication content. The required amount of communication needed decreases in terms of number of turns taken, token counts of lexical items as time passes.

3- The communication system evolves from a state to support the production of only basic lexical items having the properties of interjections, communicative actions of directive speech acts, which are basically expression of demands, emotions, attention-demanding exclamations, affirmatives or negative words. This system evolves into a state of higher levels of production of information-rich assertive (constative/declarative) communicative actions expressed by full sentence-like complex utterances.

4- A quantitative manipulation of the task structure may create temporary disturbance in the characteristics of the system in terms of changing temporarily the quantitative features of the system. That is, increasing the amount of targets creates an increase in the amount of the communicative actions required. The qualitative effect is the emergence of new coordination scripts (task sharing, task sequencing) and corresponding communication content to coordinate handling the increased number of tasks. However, basic types and ratios in the system proved to be resilient against this manipulation.

This increase in the quantitatives concerning the amount of communication should not merely be attributed to the increase in the number of target objects. There is a process of re-adapting to a new situation. New communicative conventions are to be developed to share the task that now comprises two targets. The targets may be shared, the scripts for finding one first together and then looking for the second, or searching the two targets simultaneously by separate tours of the participants must be negotiated. The negotiation not only requires some rudimentary communication about the proposed action order but also silent coordination of non-communicative actions.

This process comprises the re-adaptation activity; until the re-adaptation is achieved the coordination will be less effective and the communication will be bulky or prolix. The high values for several parameters in the blue segment in the generalized/abstract graph below depict this disturbance and the re-adaptation. Then after settlement and new coordination is achieved the difference between the horizontal orange bars (exp5-6-7-8) and the horizontal green bars(exp1-2-3) is the real difference on the qualitative parameters due to the increase in the number of targets from one to two.

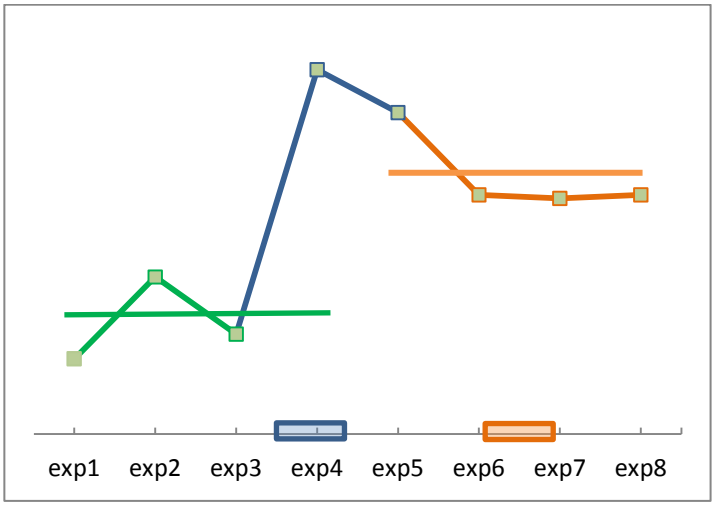


Figure 72 Generalized graph depicting the typical development of communication systems (reproduced here for convenience)

A manipulation of interrupting experimental sessions for a long period did not cause decay in the acquired skills of using the emerged communication system. The length of the break is substantial when compared to the length of the period the dyad interacted in the environment before the break (six to eight times longer, around two months *vs*. one or one and a half week). The previously developed lexical inventory was almost instantly available after the break. The negotiated strategies and acquired skills of communication specific to the experimental

constrains, as well as the lexical item retrieval performance appear to be a part of procedural memory and hence shows resilience against the manipulation.

5.1.1. THE SIGNIFICANCE OF THE RESULTS

The above results can be described as a general characterization of emergence/re-emergence of language and communication and be tested with different experimental settings of future studies. Actually there are numerous studies allowing the use of NL but those presented a special task structure and environment where the existing, shared NL had to be used in a fine-tuned way. The fine-tuning process of NL and the emergence of NL-like communication systems in the present study appear to have isomorphic features. The study by Clark and Wilkes-Gibbs, (1986) shows similar results to our results (1) and (2) presented above. (Note: Clark and Wilkes-Gibbs study involves two physically isolated participants in verbal (NL) communication about a set of (12) complex tangram figures where defining them or referring to them with the existing NL lexical items is not possible. They are given same set of figures and the task requires organizing the figures given in a random order on each table of participants, to an identical sequence on both of the tables. For a new trial same set of figures are shuffled.

In their study, even NL usage was allowed; however, a new lexicon (set of definitions, where each definition gradually gets shortened to a shortcut word, in their study) was necessary as a requirement of the given tasks' nature and the experimental environment. As a consequence, the above-mentioned general characteristics may also be tested to measure their validity in order to shed light on real world situations where the communication is constrained or the existing NL cannot handle the requirements of a new adaptive challenge.

5.2. THE DIMENSIONS OF COMPRESSION

The present study creates a case where existing communicative skills and NL competence are projected onto a communication channel that consists of a limited set of building blocks of language and a written communication medium which is limited in comparison with real embodied interaction. This projection can also be termed" compression" because a full communicative and NL system including a huge lexicon, is "squeezed" into a much smaller expressive system, in particular in terms of lexical size. Then the question is what is compressed into the emerging communication system and what is not?

Every capacity of communication and language is not attempted to be compressed into the new system. The rule of economy of the relevance theory for comprehension can be applied here for the development: only the very easily re-erectable features (word order but not tense) or highly necessary features (NP structures) are selected. This selection also shows that emergent or re-emerging communication systems are embedded in the coordination of behavioral processes (Steels, 2006a), and this embeddedness explains the power of the rudimentary sign system sufficient to facilitate the communication in our study. It is also one of the reasons why some aspects of NL or communication skills are not attempted to be re-realized through the development of the communication systems. This is because *the partial compression of NL was more than sufficient*. Similar to the partial sign system that efficiently designated the location of the target room in Galantucci's experiment, one of the dyads was only reducing the possible target locations to half by the signals they used; however, this information combined with the well-coordinated action provided a very effective task accomplishment in comparison with other dyads which had more complete representation systems. In the present study, the successful comprehension of underdetermined speaker meanings or the excessively minimalistic communication system of Couple 7 are good examples of this embeddedness. When people are able to integrate the

communicative coordination with the behavioral coordination, this action/communicative action complex turns into the problem solving strategy for the task accomplishment. Consequently there is no need for the participants to bother themselves with the revival of a full NL in a highly constrained communicative channel. This conclusion seems to be attributing a limiting role to the task structure and the experimental environment, as if they were limiting the potentials of the emerging communication system to converge on NL-like power, due to the simplicity of their demands, i.e., because the interactants do not need that much. It is quite the opposite according to the findings of the present study and according to the studies adopting a communication/joint action perspective. The task structure provides a context, a base for collaboration and the environment provides a common ground for the coordinated actions – these are the essential parts for the development of the communication system. They narrow down the expectancy space for the possible meaning/context pairs and limit efforts to comprehend the correct meaning of a rudimentary signal among the possible few, being under-determined by this very signal. Therefore, for the case of emergence of a communication system, the both a limiting and a crafting role is attributed to the task structure and the experimental environment.

It may be a good control study for this joint action/communication paradigm, to use the communication constraints of the present study and observe participants initiating developing communication in a dis-embodied context – no shared virtual world, and only the instant messaging software and the same *qwertyuıop* characters available. A well-designed but less specific, open-ended experimental task may possibly prove that, available communication tools being equal, the lack of alignment providing common goals and experimental environment, the development of a communication system will be slower in every respect. For example, a task which aims at the communication itself being the goal, might asks the participants to communicate with their anonymous partner using the

instant messaging program with given characters and symbols, with the aim to initiate a conversation and to maximize turn success ratio. It may be argued that the difficulty level of the task is higher than the one in the present study and the success will be minimal, especially in the earlier phases.

5.3. WITH OR WITHOUT A DEDICATED COMMUNICATION CHANNEL: DOES IT REALLY MATTER?

The logic behind blocking all possible communication or severely limiting it in communication game experiments is to create genuine conditions to observe the emergence of a human communication system or the mere communicative action, without the interference of existing ones (Galantucci, 2005; Scott-Phillips, Kirby, and Ritchie, 2009). In the present study, the participants' oral or written communication with NL is also severely constrained. Galantucci's study blocks the use of NL completely and gives a highly incapacitated graphical communication medium where written communication and communication of prototypical graphical signals is virtually impossible. The Scott-Phillips, Kirby, and Ritchie study designs an *embodied communication game* where the participants are just able to see the each other's action in the virtual environment but do not have any access to any communication channel at all.

However, considering the results of the present study we can say that in all three (Galantucci 2005 and Scott-Phillips 2009), the subjects managed to exploit their existing established communication and linguistic capacities in to the service of the emergence/re-emergence of a "human communication system". Galantucci provided a graphical medium – the moving pad disabled drawing characters or known shapes, only simple line segments and curves were possible. Under these constraints a novel communication system emerged, but the sign systems were based on basic signals indicating map-based location or geometrical object properties or numeration-based naming of the locations. The participants utilized

shared knowledge, notions and representation strategies of numbers, map-based location representations and features of geometrical objects.

In the Scott-Phillips, Kirby, and Ritchie study, since the only action was wandering in the 4 room area where participants had continuous visual access to each other's movements, subjects dedicated some of the movements and action time to signal their intended simple meanings by means of specific patterns of movements among the rooms. They could see each other's avatars continuously from a bird's eye view, so they made unusual/salient movements/walks within the 4-room environment. Circular prolonged moves, C shape moves, and oscillatory moves (back and forth between two rooms) were demonstrated in Figure 3 in the literature chapter. All these salient actions also converged on some shared, previously acquired graphical patterns or action patterns to maintain their salience and distinctiveness as communicative signals. They also relied on the inference on the side of the other partner that *any salient unusual distinctive action, movement pattern which is presented to me under the assumption of my presence or continuous visual access, should have some communicative intention directed at me* (see the *communicative principle of relevance* in the previous chapter, which axiomatizes each signal's presumption of own relevance). So in addition to the meaning of the signals, the question of *how these signals signal the fact that they are signals, when there is no pre-established system of communication and communication channel* receives its answer. If the participants are previously communicators of any other communication system, under the presence of each other and the continuous visual contact they have the covert knowledge of the fact that ostensive/salient behaviour implies relevance, *signalhood*, and that this is also assumed by the receiver. So that receiver invests her/his cognitive recourses to process the relevance of the signal.

The conclusion is that, under the condition that previous users of any human communication system participate, there are numerous aspects to be

considered in an emergence of communication study as well as the existence or absence of a ready-made communication channel. The assumed *theory of mind module* or *intention detection sub-module* are always available, they make the process of developing or initiating the protocols to infer the communicative intention in certain actions redundant – it is already there, within a framework of collaborative action where ostensive behaviours are habitually tracked and detected. Secondly, although the participants are deprived of any channel of communication they still have the shared representations and knowledge of the world that may be relevant to their context if they were able to communicate, so the limitation or removal of the communication channel just limits the complexity level of utilization of these representations. If the use of NL is prohibited and graphical communication is incapacitated, and if embodied actions are only the carrier of information, the complexity level and the level of sophistication of external signs will be low. The level of details that can be revealed by actions is the limit to signal previously shared representations.

5.4. FROM ACTION TO COMMUNICATIVE ACTION

The previous section on the existence of communication channels or developing ones was limited to the situations of agents' conscious and immediate need and intention to exhibit communicative behaviour and send signals. However, there exists the phenomenon that certain embodied actions, persistent patterns of bodily, sensory-motor interactions turn into communicative behaviour as a result of incremental increase in their influence on the flow of action of the observer.

In the present study the participants used their visually observable actions to coordinate their collaboration. One of these is searching for the partner as well as the target objects, usually to lead her/him to the already found target object. When one target is found the finder informs the other about the fortunate event and simultaneously searches for the partner in the environment. Since they could

not develop individual or shared mental maps of this complex environment – a dense forest of huge abstract objects – the best strategy is not trying to signal the location but finding the participant and leading her/him to the targets' location by keeping visual contact. In the case when the finder finds the partner (or *vice versa*) s/he may use some established exclamations or order to make the other participant stop and try to see/find the finder. If this signal is interpreted correctly the participant makes a 360 degree turn to scan and to try to see the finder around, but since the visual angle is very narrow in the screen-based virtual environments it is not practical. So the main solution is entering the visual path by crossing the path of the other. This is a common action pattern: the finder-searcher crosses the path, turns her/his body to the other participant and stops. This action pattern is generally interpreted as "stop what you are doing/where you are going and follow me" (note the resemblance between the Quinn 2001 study's findings presented in section 2.4.2). This action pattern is also accompanied by an utterance of "follow me" (nearly all of the couples has the corresponding lexical items and speech acts for the "follow me/I will follow you" signals). However, for establishing visual contact and the leading/following formation this visual/physical interaction is the essential action. Here, my opinion is that this pattern of interaction is ritualized after some persistent use and is in fact developed into a communicative action. It can be claimed that this is a kind of *silent coordination* (Galantucci, 2005) that employs the inferential mechanisms of the dyad members, but when this interaction pattern is implemented many times, it can be claimed that this action of path-crossing is interpreted by similar mechanisms that are also at works in the "follow me" message. Maybe at this point the broader, causal definition of communication which may have been developed for encompassing animal communication or the kind of communication systems that lack intentionality is also critical to understand this kind of modalities of emergence of human communication (see Oliphant and Nolfi's accounts in 2.4.1 and 2.4.2). Galantucci's

silent coordination and the concept of integrating the communication with the action may be particularized in theorizing some of the phenomena in the case of forming ritualized interaction patterns that has an effect of "influencing the sensorimotor flow of agents, enhance the adaptive ability of the group as a whole" (Nolfi, 2005, p. 234).

5.5. PARTICIPANTS' SUBJECTIVE COGNITIVE ACCOUNTS OF THEIR COMMUNICATIVE ACTIVITY

During the post-experimental interview (and rarely while supervising the experimental sessions) the participants are asked to provide explanations for their interpretations of the partner's messages. For example, after the first target is shared by P2 as "yq" ("yeşil küre" (green sphere)) P2 is asked about P1's response "pp":

(12:36:33 PM) partcp1: p p->?
(12:36:50 PM) partcp2: yq
(12:36:56 PM) partcp1: pp

"What does p1 means when saying "purple pyramid" with 'pp'? 'The target is purple pyramid' or 'Let's go to purple pyramid first' or 'I am going to purple pyramid'?"

Since this is uttered after the target-sharing stage in the initial phases of the session, it may not mean the assertive function "the target is pp" – it is more like a Commissive or Directive. And there is little chance of P2 to miss that interpretation. The participants report their interpretation and comprehended meaning as their immediate interpretation even in the very rare cases of misinterpretation; they do not report that they consciously evaluate a set of possible interpretations and do not rule out the less likely or irrelevant ones one by one. They do not report a relatively close or different alternative interpretation that they also negotiated individually and then ruled out with hesitation. If they report

an inferential mechanism about the interpretation of an utterance in a certain way, like, “she cannot mean this by that since we meant that with this in the previous session” the experimenter asks the question “Are you drawing these conclusions now in the post-experimental interview or did you think exactly the same way during the time of the turn taking?” In most of the cases, the participant admits that s/he was formulating the logic when an explanation was asked. These swift and heuristic conclusions can be accounted for by the built-in relevance-theoretic principles of the theory of mind module of the human cognitive system. This is not ignoring the local instances of conscious and logical problem solving to infer the meaning but the problem of inference is narrowed down to a very easy selection among a few alternatives of speech acts or word meanings.

The post-experimental interview itself may have interfered with these mechanisms of automated handling of intention detection or inferences. This may be a limitation of the study – the conversation history of each session is being rehearsed by the experimenter and the participant as an unintended consequence of the requirement of detecting the turn success. The participants are asked to report their intended meanings of messages sent and comprehensions of the messages received. And sometimes they are asked to provide an account of their choices in conversation and of their general strategies to express their meanings under the experimental constraints. The only possible option was to make use of a verbal protocol, where participants had to speak out/think aloud their own written utterances and meanings, and their comprehensions of utterances of the other dyad member in real time, synchronized with the turn-takings and recording this and then encoding them into experimental data. This design also carries the communicative interaction to a new modality and may interfere even more with the real time cognitive mechanisms and pragmatic functions during the communication.

5.6. LIMITATIONS OF THE STUDY:

Effect of experimental design preferences:

There were several design options that might have affected the content of lexical inventories and other features of the developing communication systems differently. These could have been tested in the pilot experiments phase but practical reasons (limited resources of space, time and participants) made these examinations impossible. Different experimental design options and their significance could be exemplified as follows:

i- The number and configuration of given characters ([qwertyuıop] or [asdfghjkL], [zxqti]) may prime different NLs: the first configuration which was used in this study seems to be priming English words for the all METU student participants; The latter configuration with less numbers of characters may provide a hard to initialize but – once in place – a very minimalistic communication system which might reveal some other aspects of emerging communication systems.

ii- The rules of lexical item formation: Limiting the length of lexical item by a rule such as: “if you concatenate 3 letters, a space is obligatory” or making single character lexical items obligatory.

The provided alternative lexical building blocks and word creation limitations might have created an alternative set of communication systems with which we can corroborate or question the findings based on the communication systems in this study. Indeed, some of the dyads self-imposed some of those alternative rules onto their own communication system, thus making it stricter. The instructions of the present study do not forbid or promote multi character NL-like word creation, but some couples assumed them stricter than they actually are and created possible alternative minimalistic systems, as in the case of Couple 7. Some other couples also

tried to communicate under these self-imposed stricter rules in the earlier sessions and quit only later.

The main focus of the research was not the task performance of the dyads or individual participants – the main concern was to observe the emergence and development of the communication systems in the context of a collaboration need, created by the task given to the dyads. In previous similar studies there was always a score of efficiency in task performance and this score was used to evaluate the effect and function of the emerged communication system. Here, we used the quantitative features devised to examine the development of the communication system as efficiency and maturation criteria, like type/token number of lexical items ratio, or the use ratio of assertive speech acts. These other studies in the literature (Galantucci, 2005; Scott-Phillips et al., 2009) used a general score of success that is defined independently from the features of the communication system and which was used to compare the efficiency of these systems. They had also devised methods to rule out the effect of non-communicative coordination in the overall success. This was possible by only designing very minimalistic experimental environments that allow to strictly limiting the number of locations, possible sets of movements/actions. This allowed a calculation of success possibility by chance and the increase in the success scores were attributed to the use of the emerging communication system by the dyad. Since the interaction, navigation and the environment was rich in the present study, a measure of communication-less task performance could not possibly have been developed.

The post-experimental interviews after each session may have interfered with the experience and the interaction of the participants during the experiment. Due to this interference, general subjective evaluations of the communication system and its development were probed in the last post-experimental interview or in the later sessions (Questions like “What do you think about how you developed this system?”, “How did you arrive at that comprehension of the partner’s message,

but not this one, what is your explanation?"). Alternative methods for tracking turn success and lexical item meaning might be thought of but some kind of interference seems unavoidable.

5.7. CONCLUSION

Within the ActiveWorlds virtual environment, given a collaborative task of finding targets, communication systems developed among human dyads, provided with a seriously constrained communication device, over the course of eight sequential experimental sessions. A lexical inventory evolved usually by a strategy of compressing possible NL words into the system within the limits of the dedicated communication channel.

Certain resilient patterns of development are observed in all of the participating dyads, independent from their specific strategy of developing a new system under the given constraints. These patterns can be related to a general cognitive trait of [striving for] *disambiguation* where interacting individuals display a joint effort at resolving ambiguity of meaning in the initial stages and adhere to the *cognitive/communicative principles of relevance* when there is a basic working system in the later stages. The opposing, decreasing trends of Directive and increasing trends of Assertive speech acts correspond to this characterization [striving for disambiguation]. The above-mentioned and other general characteristic strategies presented in this study may also be utilized as descriptive criteria (or may be tested for their validity) in real-world situations where communication is constrained or the existing NL cannot handle the requirements of a new adaptive challenge.

The given experimental constraints on the communication channel, the virtual environment and task structure are observed to promote the development of the communicative systems. Possessing already the general skills of communication, the members of the dyads were able to take the common

constrains they shared with their partners into consideration and produce and interpret the communicative behaviour [of the other] accordingly. As a result, the common constraints as well as the task requirements function as the ground for joint action and the emerged communication systems were an integrated part of the solution complex, i.e. the collaborative effort for the solution of the problem, namely the attainment of the experimental tasks.

This study exposes that current and past studies on the emergence of communication sharing the extended cognition, distributed cognition and joint action paradigms seem to be in dialogue and converge in terms of terminology, discussion of common theoretical problems and research agenda. However there is a lack of communication between these and the pragmatic/linguistic accounts of communication and emergence of communication, which attempt to offer a cognitive architecture account of the cognition and communication problem, by presenting a dedicated sub-module of comprehension which includes intention detection and working principles of relevance theory. This study attempted to integrate the *cognitive/communicative principles of relevance* to the explanatory scheme of the joint action studies on the emergence of communication.

REFERENCES

Akmajian, A., Demers, R., & Harnish , R. (2001). *Linguistics: An Introduction to Language and Communication.* The MIT Press.

Arbib, M. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences, 28(2)*, 105-124.

Arbib, M. (2006). *Action to language via the mirror neuron system.* Cambridge University Press.

Austin, J. (1962). *How to do things with words.* Oxford University Press.

Axelrod, R. (1981). The evolution of cooperation. *Science, 211*, 1390-1396.

Bach, K. (2005). Speech Acts. Routledge Encyclopedia of Philosophy . Retrieved from http://online.sfsu.edu/kbach/spchacts.html

Baker, M. (2003). *Lexical Categories: Verbs, Nouns and Adjectives.* Cambridge University Press.

Baronchelli, A., Felici, M., Caglioti, E., Loreto, V., & Steels, L. (2006). Sharp Transition towards Shared Vocabularies in Multi-Agent Systems. *Journal of Statistical Mechanics, P06014*.

Cichy, R. (2005). Dennettian Concept of Intentionality: Past and Present. Germany.

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

de Smith, M. (2013). Retrieved January 5, 2013, from STATSREF: Statistical Analysis Handbook: http://www.statsref.com/HTML/index.html?shapiro_wilk.html

Eryılmaz, K. (2011). Bootstrapping Shared Vocabulary in a Population - Weighted Lists with Probabilistic Choice. Ankara, Turkey.

Floreano, D., & Mitri, S. (2007). Evolutionary Conditions for Emergence of Communication in Robots. *Current Biology, 17*, 514-519.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science, 29*, 737-767.

Galantucci, B., & Sebanz, N. (2009). Joint action: Current perspectives. *Topics in Cognitive Science, 1(2)*, 255-259.

Gerard, S., & Pickering, M. (2004). Why is Conversation so easy? *Trends in Cognitive Science, 8(1)*, 8-11.

Glass, G., Peckham, P., & Sanders, J. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of. *American Educational Research Association, 42*(3), 237-288.

Goldin-Meadow, S. (2004). *The Resilience of Language. What Gesture Creation in Deaf Children Can Tell Us About How All Children Learn Language.* (J. Wellman, & H. Werker , Eds.) New York: Psychology Press.

Grice, P. (1975). Logic and conversation. (P. Cole, & J. Morgan, Eds.) *Syntax and semantics:Speech Acts, 3*, 41-58.

Grice, P. (1989). *Studies in the Way of Words.* Harvard University Press.

Hasson, U., Ghazanfar, A., Garrod, S., Keysers, C., & Galantucci, B. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in Cognitive Science*, 1-8.

Hauser, M. (1996). *The Evolution of Communication .* MIT Press.

Herrmann, E., Call, J., Hernández-Lloreda, M., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317*, 1360-1366.

Horn, R. (n.d.). Understanding Repeated-Measures ANOVA. Retrieved 12 5, 2012, from http://oak.ucc.nau.edu/rh232/courses/EPS625/Handouts/RM-ANOVA/Understanding%20Repeated-Measures%20ANOVA.pdf

Hutchins, E. (1995). *Cognition in the wild.* The MIT Press .

Hutchins, E., & Johnson, J. (2009). Modelling the Emergence of Language as an Embodied Collective Cognitive Activity. *Topics inn Cognitive Science*, 523-544.

Kaplan, F. (2005). Simple models of distributed co-ordination. *Connection Science, 17*(3-4), 249-270.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight (Ed.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form* (pp. 303-323). Cambridge.

Kirby, S. (2007). The evolution of language. *Oxford Handbook of Evolutionary Psychology*, 669-681.

Kirby, S. (n.d.). Natural language from artificial life. *Artificial Life, 8*(2), 185-215.

Lipson, H. (2007). Evolutionary Robotics: Emergence of Communication. *Current Biology, 17*(9), R330-R332.

Melinda, C., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). *Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age.* University of Chicago Press.

Nolfi, S. (2005). Emergence of communication in embodied agents: co-adapting communicative behaviours. *Connection Science*, 231 - 248.

Noveck, I., & Sperberger, D. (2007). The why and how of experimental pragmatics: The case of 'scalar inferences'. (N. Burton-Roberts, Ed.) *Advances in Pragmatics*.

Oliphant, M. (1997). Formal Approaches to Innate and Learned Communication: Laying the Foundation for Language. San Diego, USA.

Origgi, G., & Sperberger, D. (2000). Evolution, communication and the proper function of language. In P. Carruthers , & A. Chamberlain (Eds.), *Evolution and the Human Mind: Language, Modularity and Social Cognition* (pp. 140-169). Cambridge: Cambridge University Press.

Quinn, M. (2001). Evolving communication without dedicated communication channels. *Advances in Artificial Life: Sixth European Conference on Artificial Life (ECAL 2001), 2159*, 357-366.

Radford, A. (1997). *Syntax: A Minimalist Introduction .* Cambridge: Cambridge University Press.

Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neuroscience, 21*(5), 188-194.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, 1926-1928.

Salomon, G. (1993). *Distributed Cognitions: Psychological and Educational. Considerations.* Cambridge: Cambridge University Press.

Schegloff, E., Jefferson, G., & Sacks, H. (1977). The Preference for Self- Correction in the Organization of Repair in Conversation. *Language, 53*(2), 361-382.

Schwab, A. (n.d.). Repeated Measures ANOVA Problems. Retrieved 12 8, 2012, from http://www.utexas.edu/courses/schwab/sw388r7_spring_2007/SolvingProblemsInSPSS/Solving%20Repeated%20Measures%20ANOVA%20Problems.pdf

Scott-Phillips, T., Kirby, S., & Ritchie, G. (2009). Signalling Signalhood and the Emergence of Communication. *Cognition, 113*, 226-233.

Searle, J. (1969). *Speech acts: an essay in the philosophy of language .* Cambridge: Cambridge University Press.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences, 10*(2), 70-76.

Shintel, H., & Keysar, B. (2009). Less is more: A minimalist account of joint action in communication. *Topics in Cognitive Science,, 1*, 260-273.

Shockley, K., Richardson, D., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science, 1*(2), 305–319.

Sotillo, S. (2000). Discourse Functions and Syntactic Complexity in Synchronous and Asynchronous Communication. *Language Learning & Technology, 4*(1), 82-119.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition.* Oxford: Blackwell.

Sperberger, D., & Wilson, D. (1986). *Relevance: Communication and cognition .* Oxford: Blackwell.

Sperberger, D., & Wilson, D. (2006). Pragmatics. (F. Jackson, & M. Smith, Eds.) *Oxford Handbook of Philosophy of Language*, 468-501.

Steels, L. (2003). Evolving Grounded Communication for Robots. *Trends in Cognitive Sciences, 7*(7), 308-312.

Steels, L. (2003). Language re-entrance and the 'Inner Voice'. *Journal of Consciousness Studies, 10*(4-5), 173–185.

Steels, L. (2006). Semiotic Dynamics for Embodied Agents. *IEEE Intelligent Systems, 21*(3), 32-38.

Steels, L. (2006a). Experiments on the emergence of human communication. *Trends in Cognitive Sciences, 10*(8), 347-349.

Steels, L. (2007b). Language Originated in Social Brains. In O. Vilarroya, & F. Forn-Argimon (Eds.), *Social Brain Matters: Stances of Neurobiology of Social Cognition* (pp. 223-242). Amsterdam.

Steels, L., & McIntyre, A. (1999). Spatially Distributed Naming Games. *Advances in complex systems, 1*(4), 301-323.

Steels, L., & Kaplan, F. (1999). Collective learning and semiotic dynamics. In D. Floreano, J. Nicoud, & F. Mondada (Ed.). (pp. 679-688). Berlin:Springer-Verlag: ECAL99.

Stenning, K., Lascarides, A., & Calder, J. (2006). *Introduction to Cognition and Communication .* The MIT Press.

Thagard, P. (2005). *Introduction to Cognitive Science.* The MIT Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions The origins of cultural cognition. *Behavioral and Brain Sciences, 28*(5), 675-691.

Tsai, C., Kuo, W., Jing, J., Hung, D., & Tzeng, O. (2006). A common coding framework in self-other interaction: evidence from joint action task. *Experimental Brain Research, 175*, 353-362.

Vogt, W. (1999). *Dictionary of Statistics and Methodology: A Non-Technical Guide for the Social Sciences (2nd ed.) .* Sage Publications.

Wagner, K., Reggia, J., Uriagereka,, J., & Wilkinson, J. (2003). Progress in the Simulation of Emergent Communication and Language. *International Society for Adaptive Behavior, 11*(1), 37-69.

Wilson, D., & Keil, F. (1999). *MIT Encyclopedia of the Cognitive Sciences.* The MIT Press.

Wilson, D., & S. D. (2002). Pragmatics, Modularity and Mind Reading. *Mind and Language, 17*, 3-23.

Wilson, D., & Sperberger, D. (2004). Relevance Theory. (L. Horn, & G. Ward, Eds.) *The Handbook of Pragmatics*, 607-632.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *PNAS, 105*(13), 5012-5015.

APPENDICES

APPENDIX A PRESENTATION AND INSTRUCTIONS OF THE EXPERIMENTS

Deney Tanıtımı ve Kuralları

1- Bu deney, katılımcıların bilgisayar tabanlı sanal gerçeklik ortamında kendilerine verilen ortaklaşa görevleri beraberce başarmaya çalışmalarına ve iletişimsel eylemlerde bulunmalarına dayalıdır.

2- Deneyde katılımcıların bireysel başarıları veya katılımcı çiftin başarısı ölçülmemektedir. Gözlemlenmesi beklenen süreçler içinde, iletişim ve görevi gerçekleştirme çabası boyunca yaşanan sorunlar, çıkmazlar ve hatalar da önemli ve değerli veri kaynaklarıdır. Katılımcı kimliğini şahsen tespit edecek kişisel bilgi toplanmayacak ve veri analizinde kullanılmayacaktır.

3- Deney, katılımcı çifte verilen kartlarda yazılı görevi, deney alanında dolaşarak gerçekleştirmeye dayanmaktadır. Katılımcıların her biri, bu ortaklaşa görevi gerçekleştirmek için gerek duyarsa, kendisine sağlanmış online yazışma programından yazılı işaretler gönderebilir. Bu işaretler sadece klavyenin en üst satırındaki

Q, W, E, R, T, Y, U, I, O, P harfleri, **?** ve -> işaretleridir.

(ok işareti -> şeklinde – ve > (klavyenin sol alt kösesindeki) büyüktür işareti ile oluşturulabilir, küçük veya büyük harf kullanılabilir)

*4- Verilen görevler genellikle bir yeri ve nesneyi bulmak şeklinde olacaktır. Hedefin deneklerden sadece birine verildiği durumlarda, hedef bulunduğunda diğer deneğin amaçlanan **hedefin ne olduğunu tam olarak anlamış olması beklenmektedir**.*

5- Her biri 10 dakika civarında sürmesi beklenen 6 oturumun iki farklı günde tamamlanması planlanmaktadır. Oturum bittikten sonra katılımcıların her biri ile deney sonrası mülakatı yapılarak, deney sırasındaki iletişimsel eylemleri ile ilgili bilgi toplanacaktır.

6- Katılımcıların deney oturumlarının tamamı bitmeden, deney esnasındaki iletişim kurma şekilleri ve genel olarak deney hakkında birbirleriyle konuşmaması sağlıklı veri toplanması ve araştırmanın temel mantığı açısından çok kritiktir, bu konuda hassasiyet gösterilmesini önemle rica ediyoruz.

APPENDIX B EXPERIMENT TASKS FOR THE 8 EXPERIMENTS

	PARTICIPANT 1	PARTICIPANT 2	
EXP1	**Blue Sphere**	--	SINGLE TARGET
EXP2	**Green Cube**	--	SINGLE TARGET
EXP3	--	**Yellow Pyramid**	SINGLE TARGET
EXP4	**Green Sphere**	**Green Cube**	DOUBLE TARGET
EXP5	**Green Pyramid**	**Brown Sphere**	DOUBLE TARGET
EXP6	**Purple Cube**	**Purple Pyramid**	DOUBLE TARGET

TWO MONTHS GAP

EXP7	**Purple Sphere**	**Green Cube**	DOUBLE TARGET
EXP8	**Green Sphere**	**Green Pyramid**	DOUBLE TARGET

Experiment 1:

P1: Mavi küreyi bul, diğer katılımcıyı kürenin yanına götür. (Find the **blue sphere** and take the other participant near to it)

P2: Görevi diğer katılımcıya söyledik. (The task is given to the other participant)

Experiment 2:

P1: Yeşil küpü bul, diğer katılımcıyı küpün yanına götür. (Find the **green cube** and take the other participant near to it)

P2: Görevi diğer katılımcıya söyledik. (The task is given to the other participant)

Experiment 3:

P1: Görevi diğer katılımcıya söyledik. (The task is given to the other participant)

P2: Sarı piramiti bul, diğer katılımcıyı piramitin yanına götür. (Find the **yellow pyramid** and take the other participant near to it)

Experiment 4:

P1: Yeşil küreyi bul; diğer katılımcıyı kürenin yanına götür. (Find the **green sphere** and take the other participant near to it)

P2: Yeşil küpü bul; diğer katılımcıyı küpün yanına götür. (Find the **green cube** and take the other participant near to it)

Experiment 5:

P1: Yeşil piramiti bul; diğer katılımcıyı piramitin yanına götür. (Find the **green pyramid** and take the other participant near to it)

P2: Kahverengi küreyi bul; diğer katılımcıyı kürenin yanına götür. (Find the **brown sphere** and take the other participant near to it)

Experiment 6:

P1: Eflatun küpü bul; diğer katılımcıyı küpün yanına götür. (Find the **purple cube** and take the other participant near to it)

P2: Eflatun piramiti bul; diğer katılımcıyı piramitin yanına götür. (Find the **purple pyramid** and take the other participant near to it)

Experiment 7: P1: Eflatun küreyi bul; diğer katılımcıyı kürenin yanına götür. (Find **purple sphere** and take the other participant near to it)

P2: Yeşil küpü bul; diğer katılımcıyı küpün yanına götür. (Find the **green cube** and take the other participant near to it)

Experiment 8:

P1: Yeşil küreyi bul; diğer katılımcıyı kürenin yanına götür. (Find the **green sphere** and take the other participant near to it)

P2: Yeşil piramiti bul; diğer katılımcıyı piramitin yanına götür. (Find the **green pyramid** and take the other participant near to it)

Notes: Some of the targets are recurrent but they are not usually the same ones in terms of location; the experimental environment minimally modified between sessions. If necessary, the general structure of the maze is preserved, only colors of the target objects are changed to the ones in the instructions for that session.

For example the green cube in the second experiment which was given as the target to participant 1 is not the same one with the green cube in the fourth one given to participant 2. Before the experiment 4 leaving all the remaining objects intact, the color of the green cube the second experiment is changed to red and another cube in a different area is modified to a green color. So the participants did not trade the effect of familiarity with the object location with the communicative effort.

But other recurrent targets exist in more than one experiment, where the ones given before and after the two months break. The break was before the last two experiments (namely these targets recurred in 7^{th} and 8^{th} experiments). It is very obvious that participants could not develop a familiarity of objects and their colors/locations in the consecutive sessions, they were not

even sure that they were in the same environment, for the first 4 or 5 sessions, because in each session their starting point was changed and the dimensions and the dispersion of the objects did not allow them to see the environment with a panoramic view. The only general information that they may had acquired was there was a predominantly brown objects region, a predominantly red objects region and a predominantly light blue/turquoise objects region. The targets were usually in these regions with different colors. So you cannot take advantage of going to brown region the find a target like a "brown sphere". For these reasons, the green pyramid, the green sphere and the green cube in the targets of experiments 7 and 8 were the same targets with the targets in experiment 5, experiment 4 and experiment 2 respectively. The participants reported that, when they were given as targets, they did not recall the locations of these targets; even some reported that they did not remember that they had such a target. "Wasn't it a purple pyramid in the previous experiment?" (About the target for P2 in the 8th experiment, speaks a participant, which is given both in the 5th and 8th after a 3 sessions and 2 months period.

APPENDIX C GUIDELINE FOR SPEECH ACT CATEGORY JUDGMENTS

Inter –Rater Reliability Analysis of Speech Act Categories Judgments

1- Speech Act Categories: Sources: Searle, Austin, K. Bach and R.M. Harnish

	Definition of Speech Acts
Assertive	**affirming**, alleging, announcing, answering, attributing, claiming, classifying, concurring, confirming, conjecturing, denying, disagreeing, disclosing, disputing, identifying, informing, insisting, predicting, ranking, reporting, **stating,** stipulating (also categorized as **Constative**)
Directive	**advising**, admonishing, **asking**, begging, dismissing, excusing, **forbidding**, instructing, **ordering**, permitting, **requesting**, requiring, suggesting, urging, warning.
Expressive	apologizing, condoling, **congratulating**, **greeting**, thanking, **accepting** (acknowledging an acknowledgment) (also categorized as **Acknowledgment**)
Commissive	**agreeing**, guaranteeing, **inviting**, **offering**, **promising**, swearing, volunteering

Assertive**:** (also called Constative) Stating, disagreeing, announcing and answering, affirming. Usually statements about facts, which has truth value.

Directive: Utterances that cause or intend to create an action (in terms of behavior or communicative action) of hearer, as a response. Ask someone to do (or not to do) something, asking questions (to make someone answer).

Expressive: Utterances that let the hearer know about the psychological state of speaker or that *acknowledges* the hearers mental or emotional situation,

Commissive: Utterances (commitments) about future actions which promises, offers, invites or agrees to do something.

EXAMPLES:

Yes → Commisive (accepting, agreeing on an offer or promising something) Are you coming? ("Yes, I will come") (Commisive)
Yes→Expressive (Accepting an apology, congratulation etc., also "yes, I understand" is about internal psychological condition) Oleey we found it!! "Yes, that's great" (Expressive)

Yes→Assertive (affirming a statement about external world) We found the target. – "Yes, we did" (affirming, Assertive)

2- Indirect speech acts:

These have intended meanings that are different from their literal meanings
Hearers recognize their real meaning based on the context

i-) It's cold in here (Assertive)

I want you to turn up the heat or close the door. (**Directive**, real speech act category)

ii-) Would you mind helping me carry some of my stuff? (Directive, interrogative)

Help me carry some of my stuff. (Directive, action request)

In the data analysis phase, the rater is allowed to make an inference about whether the speaker's speech act indirect or not, by relying on the speakers report on his/her intention.

3- Speech act category judgments require context evaluation, since communicative tools and patterns are limited, certain prototypical utterances that may seem to be dedicated to a certain speech act is utilized for another one. This utilization requires a mutual assumption of successful mind reading of participants.

Here are some cases of category judgments which require a deviation from the apparent speech act category of an utterance:

1- Directive → Commissive correction: Assumption about the authority position between participants is reconsidered (If there is no authority position but collaboration or vice versa)
EXAMPLE:

LOG	*partcp1 Interpretation*	*partcp2 Interpretation*	*partcp1 Intention*	*partcp2 Intention*	*context*	*S/F*	Speech Acts
(10:51:08 PM) **partcp1:** oq qo to pluto qure		ok, eflatun küreye gidelim	ok, eflatun küreye gidelim		visual contact	Success	Commissive

Here the lexical binding of words to an NL utterance requires a Directive Speech Act judgment (*order*) but the speaker reports explicitly the intention of *offering* by "gidelim" and the utterance is categorized as **Commissive** Speech Act.

2- Assertive → Commissive and Assertive → Directive correction:

The indirect speech act approach is adopted; the **information given (stated) by** the **utterance implies an offer or order** for a related action.

3- Expressive → Commissive: -"Let's go" **-"OK",** here **"OK"** is agreeing with an offer not an acknowledgment in the *expressive* sense

4- Expressive → Assertive: Expressive utterance is used to give information

5- Assertive → Expressive: Acknowledging an expressive speech act may be classified as Assertive act of affirming, in fact acknowledging the acknowledgement is by definition an expressive speech act. (HE Exp 6 line 7-8), “OK, I understand”

6- Commissive → Expressive; and Assertive → Expressive: (OK (I understand) is also Expressive)

Notes:

1- Unless there is an opposite “**Strong Contextual cue**”, or **experimenters note** in the Post experimental interview data, in Context column, Speech Act Category judgment is made according to the lexical content.

"**Trqt qree top**" "**target is green sphere**": an Assertive speech act about the targets. *We do not make an comment of* implication that it means “Let’s find this target”, or an order “Find the green sphere” it may be Commissive or Directive.

2- "**Trqt qree top** -->" If is it noted as "target is green sphere (let’s go)" it is Commissive. Depending on the note about the intention it may be Directive.

3- An affirmative/positive response to a Commissive or Directive (offer/order) speech act (like "ye", "oq" "t") are Commissive.

4- An affirmative/positive response to an Assertive utterance' ("yes" , "ok") is Assertive. But if there is a note indicating a intention of “I understood” or a cue In that direction it is categorized as Expressive.

-- "**Trqt qree top**"
-- "**Ye**"
The "ye", yes, is assertive. An affirmation with the meaning of "yes target is green sphere"
-- "**Trqt qree top**"
-- "**Y, ı qot ıt**" is expressive since it contains the expression of the mental state. Same if there is a note in that sense.

5- "Trtq oq" is assertive if the explained intention is “target is found” but if there is a special expressive interjection and intention noted in the lexical content of the utterance, the turn is categorized as expressive (In a celebrative, congratulation manner)

6- If an utterance is completed in two turns, or speaker divided the utterance in to two by stroking enter in the middle of the sentence, then that consecutive turn is accepted as a single utterance. The final speech act category judgment is applied to both of the utterances in two turns.

CURRICULUM VITAE

MURAT ULUBAY

Computer Center
Middle East Technical University
06531 Ankara, Turkey
e-mail: ulubay@gmail.com

EDUCATION

Degree	Institution	Year of Graduation
PhD	Middle East Technical University Cognitive Science/Informatics Institute	2013
MSC	Middle East Technical University Sociology	2002
BS	Middle East Technical University Mathematics Education	1997
High School	Kadıköy Anadolu Lisesi	1993

WORK EXPERIENCE

Year	Place	Enrollment
2000-2003	Middle East Technical University Computer Center	Programmer
2003-	Middle East Technical University Computer Center	Research Assistant

FOREIGN LANGUAGES

English (Advanced), German(Basic), Arabic (Basic), Persian (Basic).

TEZ FOTOKOPİ İZİN FORMU

ENSTİTÜ

Fen Bilimleri Enstitüsü ☐

Sosyal Bilimler Enstitüsü ☐

Uygulamalı Matematik Enstitüsü ☐

Enformatik Enstitüsü ☒

Deniz Bilimleri Enstitüsü ☐

YAZARIN

Soyadı : ULUBAY .....
Adı : MURAT.....
Bölümü : BİLİŞSEL BİLİMLER.....

TEZİN ADI (İngilizce) : RESILIENT FEATURES OF RE-EMERGING DYADIC COMMUNICATION SYSTEMS IN AN INTERACTIVE VIRTUAL ENVIRONMENT

TEZİN TÜRÜ : Yüksek Lisans ☐ Doktora ☒

1. Tezimin tamamı dünya çapında erişime açılsın ve kaynak gösterilmek şartıyla tezimin bir kısmı veya tamamının fotokopisi alınsın. ☒

2. Tezimin tamamı yalnızca Orta Doğu Teknik Üniversitesi kullanıcılarının erişimine açılsın. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.) ☐

3. Tezim bir (1) yıl süreyle erişime kapalı olsun. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.) ☐

Yazarın imzası ..... Tarih 21.01.2013