EVOLUTION OF COMPOSITIONALITY WITH A BAG OF WORDS SYNTAX

by

Ahmet Engin Ural

A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Electrical & Computer Engineering

Koç University

August, 2008

Koç University
Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Ahmet Engin Ural

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Assist. Prof. Deniz Yüret

Assoc. Prof. Aylin Küntay

Assist. Prof. Engin Erzin

Date:

To my grandfather, Ismail Bedenli and to my family

Dedem Ismail Bedenli'ye ve Aileme

ABSTRACT

In the last two decades, the idea of an emerging and evolving language has been studied thoroughly. The main question behind this kind of studies is how a group of humans reaches an agreement on the phonology, lexicon and syntax. The improvements in computational tools led the researchers build and test models that have been ran computer simulations to answer the question. Although the models are mere reflections of the reality, the results have been often useful and insightful. This dissertation follows the same line and proposes a new model, tested in a game based simulation methodology. Besides, this work tries to fill the gap in the studies of lexicon compositionality and proposes a plausible explanation for the transition from single word naming to multi word naming. The direction of the results is in line with the previous research such as the emergence of a stable and communicative language. Moreover compositionality in lexicon is observed with a very simple bag of words syntax. The parameters influencing the results are analyzed in depth. Even though the model does not meet the standards of the real world, future work hints insightful facts about the transition from single word naming to syntax.

ÖZETÇE

Bir dilin hiç bir öncülü olmadan ortaya çıkması ve evrim geçirmesi bilimadamlarınca son 20 yıldır detaylıca araştırıldı. Bu araştırmaların arkasındaki esas soru ise bir grup insanın nasıl olup da ortak bir ses sistemi, kelime haznesi ve dilbilgisi üzerinde uzlaşması ve bu uzlaşmaya göre iletişim kurması oldu. Gelişen hesaplama teknikleri ve bilgisayar araçları sayesinde bu soruya cevap olabilecek sistemleri bilgisayarlarında modelleyip sonuçlarını alabildiler. Bu mezuniyet tezi de bu araştırma çizgisi dahilinde yeni bir model önermektedir. Bu yeni modelin cevaplamak istediği soru ise, tek kelimelik dillerden, çok kelimeli dillere geçişin nasıl gerekleştiğidir. Sonuçlar, tutarlı ve düzenli bir dil ortaya çıkması açısından önceki sonuçları desteklemektedir. Yeni bir sonuç olarak da anlamlı parçalardan oluşan çok kelimeli ve basit bir dilbilgisine sahip olan bir dil ortaya çıkmıştır. Tezde ayrıca bahsedilen sonuca etki eden parametreler incelenmiştir. Model ve deneyler olgunlaşmamış ve basit gerçeklemeler olmalarına rağmen, geliştirilmeleri durumunda tek kelimeden dilbilgisine geçişin nasıl olduğu konusunda daha fazla bilgi ve öngörü verebilecektir.

ACKNOWLEDGMENTS

I thank to my advisor, Deniz Yüret, for sharing my enthusiasm in understanding and exploring human mind, motivating me when I need encouragement most, for 6 years, beyond the lifetime of this thesis.

I thank TUBITAK and Koç University for their financial support in my education.

I thank my roommates, Ergun Biçici and Mehmet Ali Yatbaz, for sharing a room and my interest in language.

Emrah Onur Toprak and Ahmet Gündoğan, my 'tarabya sırtları' buddies, have been my off campus supports and always backed me during the times I got alienated.

During the course of this thesis, with *Işıl* I felt rejuvenated and prepared for my work. Thanks to *Işıl* , this research is more colorful.

Lastly, this dissertation is dedicated to my family, whose support is invaluable and to my grandfather, whom I will be missing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

ILF	Iterated Learning Framework
ML	Machine Learning
MNG	Multi word Naming Game
LAD	Language Acquisition Device

Chapter 1

INTRODUCTION

Language is usually called as the hardest problem in understanding the human mind. Although humans are not the only animals on earth to use a communication system such as language, it can be claimed that the human language is far most the most complex of all. There have been many attempts to teach other animals such as apes (Gardner and Gardner, 1969; Savage-Rumbaugh, 1985) or dolphins (Pack and Herman, 2004; Heidi, 2008) language. If we compare such animals' acquisition of language, which requires a lot of effort both on the teaching and on the learning side, with a two year old human's, which occurs without explicit teaching, it will be obvious that human mind achieves it in a very completely different way.

The question of "how a human mind solves the problem of acquiring a natural language" still lacks a full, comprehensive answer, despite many explanations since the times of the ancient Greece. The answer is the beyond of the scope of the thesis, too. In spite of this frustration in science in the last five decades, many cognitive phenomena behind the language skill have been revealed. Many findings such as the identification of the critical age in language acquisition and the discoveries in bilingualism, child language, cognitive linguistics have helped us have a grasp in understanding how human mind solves the language puzzle.

The studies on language caused two main debates. First of them is nativist vs non-nativist. Second one is gradual language evolution vs jumps in language evolution. As the focus of the thesis is not addressing both discussions in detail, they will only be mentioned shortly.

The former debate can be summarized around the question of "Is language pre-wired, innate or is it by product of other cognitive skills?". The nativist fraction, initialized by Chomsky with his famous book, 'Aspects of the theory of syntax' (Chomsky, 1965), argues that a great deal of linguistic abilities of humans is innate. Chomsky claims that there is a language acquisition device (LAD) in human mind that has been evolving through ages. Others argue that general cognitive abilities are sufficient for linguistic skills to emerge and specialized LAD is not necessary (Tomasello, 1985). The debate is far from a conclusion; however it is obvious that both explanations have some useful insights, in addition to drawbacks.

Latter discussion's main domain is the path from the animal signals to the modern human language. The main obstacle before discussing this issue is that language evolution did not leave any fossils behind so that we can figure out the path it followed. Despite the lack of physical evidence we can still propose some theories and test those theories. Bickerton argues that there is a protolanguage stage before getting to fully developed human language therefore there should be a sudden jump (Bickerton, 1998). On the other hand, Jackendoff offers an explanation of a gradual trajectory of language evolution. Maynard-Smith and Szathmary argue for a gradual evolution in terms of evolutionary biology (Maynard Smith and Szathmry, 2000).

What can this thesis and similar language model simulations offer to these two crucial debates in linguistics? It should be noted that, even though the language defined in those simulations is not identical to a natural language, it is usually a plausible model of it. First of all by these plausible models, language evolution experiments give powerful insights on the requirements for a language to emerge. These requirements may be cognitive, environmental or input based. Such conclusions apparently help answering the question of innateness. Also the type of the required cognitive abilities may be an important clue, i.e. whether it is specifically language related or not. Secondly by changing the settings, such as input, meaning space, environment, we can have useful insights about the possible trajectories of language evolution. For instance, there have been many studies that offered various paths for that trajectory.

The language evolution studies can be separated based on their methodology. Some of them uses anthropologic and sociolinguistic methodologies (Davidson, 2003; Comrie and Kuteva, 2004), some use purely mathematical approaches (Nowak et al., 2000; Ferrer and Sole, 2003) and there are various computational studies to explain the trajectory of language evolution (Kirby, 2007; Kirby and Brighton, 2006). More details of the language evolution studies will be explained in the next section.

Besides the broad range of the language evolution studies, this thesis has a more specific concentration: the compositionality of the lexicon with simple bag of words syntax. This syntax contains only one predicate, which is the 'and' operator. In order to observe the concentrated field, the thesis follows a methodology of language game based models. The language game based model can be categorized specifically as a multi word naming game (MNG), although it is more generalizable than previous examples of MNG whose domain was the color. MNG simulations are discussed in Section 2 in detail. On the compositionality issue besides MNG simulations which are interested in intra generational evolution of language, Kirby (Kirby, 2007; Kirby and Brighton, 2006) and Batali (Batali, 1998a) have also some research differing from MNG simulation in both methodology and approach. In Kirby, the approach is inter generational as he uses the framework of Iterated Learning Framework, which is discussed in the next sections. In Batali, the set of meanings is predetermined and the work aims to find out if the signals get decomposed into sub parts, to infer the meaning by the neural networks.

The outline of the thesis is the following. The language evolution simulations are discussed in Chapter 2 in depth. In Chapter 3, the proposed new model is explained in detail. In Chapter 4, four experiments and their results are presented. The first of these experiments demonstrates the base result of this thesis. The other experiments modify some of the parameters to see their effect on compositionality. Lastly Chapter 5 includes the comparison of this study with the previous work and evaluates the dissertation.

Chapter 2

LANGUAGE EVOLUTION SIMULATIONS

Before computational methods were employed, the studies on language evolution relied on the intuition of their authors and were hard to falsify. Therefore there were many studies from various fields such as archeology and anthropology, and most of them were hard to be both proved and disproved. However computational approaches enabled researchers to build a model, which is supposed to be realistic and plausible, and test the model in computational simulations. Therefore the theory can be practically tested.

This section discusses the previous work in the following outline. Firstly the previous work is categorized by their focus and methodology briefly in Section 2.1. In Section 2.2, the previous work is compared and contrasted by discussing specifics of the design and approaches in the components of the simulations.

2.1 Categorization of language evolution simulations

The categorization of language evolution simulations can take into account either the focus of the study or its methodology. These two classes of the studies are briefly mentioned in the following sections. The details of their approach and design is compared and contrasted in the Section 2.2, because it is easier to compare the approaches by observing the effects of specific changes in the components of the simulations.

2.1.1 Language evolution simulations by their focus

Language evolution simulations can be classified based on their focus as the following.

- Innate LAD (Yamauchi, 2001; Turkel, 2002)
- Phonetics (de Boer, 2001, 2002; MacNeilage and Studdert-Kennedy, 1984)
- Lexicon (Hurford, 1989; Vogt, 2000; Steels and Kaplan, 1998; steels and Kaplan, 2002; Pefors, 2000)

- Concepts and prototypes (Steels, 1996; de Jong, 2000; Belpaeme, 2002; Laskowski, 2006)

- Syntax (Kirby, 2001, 1999; Briscoe, 2002)

- Compositionality (Kirby, 2007; Smith et al., 2003; Neubauer, 2004)

2.1.2 Language evolution simulations by their methodology

The main method of language evolution simulations is known as agent based modeling (Steels, 1997). This kind of modeling tries to analyze the behavior of agents which are programmed at a very fine grained level (Steels, 1997). While these agents are participating in local interactions, global properties may emerge.

In agent based simulations, in particular, the agents, which have internal cognitive capacities such as memory and learning schemes, are the common element. In some experiments those capacities do change throughout the simulation to model genetic evolution (MacLennan, 1992; Briscoe, 1997). Another common element is the environment which is defined by a meaning space, with which signals are associated. On the other hand, there are some differences in the parameters of various simulations. For instance in some simulations the population size is not fixed, i.e. there is an outflux and an influx (MacLennan, 1992; de Boer, 1997; Hutchins and Hazlehurst, 1995; Kirby, 2000, 1999). The linguistic interaction between agents is often defined differently in different experiments. In Steelstype experiments, the interaction is language games. In Batali (Batali, 1998a) and Kirby (Kirby, 2007; Kirby and Brighton, 2006), the interaction is through neural networks.

The latest publications indicated that the language evolution studies are using two main methodology, which is language game models and Iterated Learning framework.

Language game models

The model which is implemented for this dissertation is a language game models (Steels, 1996). The language games are formalized interactions between agents of the population (Looveren, 2005). The task of the agents in language games is two fold. If the agent is the speaker, its task is to describe accurately the referent ,topic, in a context which is in the environment. If the agent is the hearer, its task is to identify the referent correctly, by using the linguistic cue, signals, i.e. the utterance.

The language games initiate as the following, two agents, a hearer and a speaker, are chosen from the population. The speaker selects a topic. The type of the topic depends on the design of the simulation. In some simulations the topic is an event (Steels) or an object identified with a feature vector (Steels, 1996). Then the speaker sends a signal, or a set of signals, to describe the topic. The hearer receives the signal, interprets it and points the topic with a non linguistic cue. The speaker agrees or not, by pointing the topic which he meant to speak about. This iteration of a kind of game is repeated by many times within the population. Some researchers opt to reproduce a new set of agents, by crossing over the fit and older genotypes.

At the end of the simulation, the model is evaluated by various evaluation measures which will be discussed in the following sections.

Besides the common elements of the language game models presented above, there are some variations in those models. These variations, which is discussed in Section 2.2, are made for specific purposes e.g. to analyze different variables or to create a more realistic mode.

Iterated Learning Framework (ILF)

As Kirby and Hurford (Smith et al., 2003) proposed in their work, they aimed not only to explain phenomena such as the emergence of compositionality but also to create a framework in which language evolution simulations can be run. This framework has four main components; meaning space, signal space, learning agents, teaching agents. Meaning and signal spaces may vary based on the environment, such as random or structured. Between those two spaces, teaching agents use a language based on their hypothesis of the language. Later on learning agents are trained by the utterances of the teaching agents. After training phase, learning agents end up with their own hypothesis over the meaning-signal pairs. Then the teaching agents are removed from the simulations. This iteration is repeated for more than thousand times, so that language is stabilized. First iteration is different then the others; first users of the language i.e. first generation teachers, use random pairings.

There are a few things to be mentioned about the work of Smith et al (Smith et al., 2003). In his work there is no horizontal cultural transmission. However there are studies which use ILF and focus on intra generational cultural transmission (Batali, 1998a) or the communicative function of language (Smith, 2002).

Besides the work of Smith et al. there are also other studies that focus on compositionality and syntax. Batali showed that compositionality emerges as agents hypothesize meaning-signal pairs. In his study meanings are predetermined and discrete valued (Batali, 1998a).

2.2 General framework of the simulations

This section identifies the simulations' main components and concepts on which it compares the previous studies are based.

2.2.1 Signal space

Signal space is the form of the communication in language games. Signal space may contain one or multiple word utterances. Although some studies, that implement multiple word utterances, include a syntax, the latter section does not contain the discussion of syntax studies. Because this study do not focus on syntax. In some cases all the elements of the signal space are determined beforehand, i.e. the lexicon is given to the agents before the experiment starts.

One word

One of the first language evolution simulations which is created by Hurford, uses one word length signal space (Hurford, 1989). His experiment attempts to explain the children's language acquisition strategies. Another one word signal space example belongs to Yanco and Stein (Yanco and Stein, 1993). In their case the lexicon is determined beforehand and the robots are expected to acquire their meanings. MacLennan did comparable studies (MacLennan, 1992). Cangelosi and Parisi (Cangelosi and Parisi, 1996) analyzed the emergence of warning signals, e.g. an agent signals "it is poisonous", when it warns the others about a mushroom in the environment. In their model which does not impose the agents to signal, the genetic variability of the agents resulted in their choice of altruistic signaling. So that the emergence of communication can be monitored. In general it is observed that the early examples of the language simulations employed simple one word length lexicon. One of the reasons behind is the fact that they generally tried to address the role of the communication, rather than the role and the emergence of language in particular.

The later experiments followed the previous ones which focused on the emergence of the lexicon, so that the agents are imposed to signal and the lexicon is not determined previously. De Jong and Steels produced important examples of those experiments. In their experiments, the agents use one word length utterances to describe a topic in a context. The experiments conclude with a coherent lexicon.

Another variant of this type of experiments includes stochasticity. This is done by letting agents to misunderstand the word that they receive or by mutating their lexicon randomly (steels and Kaplan, 2002). De jong and Steels (de Jong, 2000; Steels, 1996) employ this variant differently, as sometimes the agents in their experiment do not pick the most optimal signal in their repository. The agents use a probabilistic model over all signals and picks a signal based on the signal's probability.

The language games in these experiments are referred as "simple naming game" (Looveren, 2005). They exhibit a very simple structure of communication, exchanging one signal at a time.

Multi words without syntax

The difference between the single word games and multi word games is that the signal received by the agent can be decomposed into smaller parts all of which are associated with different meanings out of the meaning space (Meaning space is described in detail in the following section.). In many studies this property is called as compositionality (Vogt, 2000). However in the literature there is an unsettled discussion about compositionality and syntax as (Looveren, 2005) points out in his PhD thesis. Many authors refer those terms interchangeably. Syntax combines compositional parts via syntactical rules. As it will be discussed in later section, multi word games utilize a very small subset of syntactical space, which is simply the 'and' predicate. This structure is an example of an uncomplicated syntax, such studies are not so useful in terms of the the explanation of syntax emergence. Thus I would prefer to differentiate this sort of research from the other syntax research and include it in the research of lexicon.

There are a few studies on multi word language games. First of them belonging to Crumpton, similar to MacLennan's work, which is discussed in the previous section (Crumpton, 1994). In his experiment the agent sends a signal, does an action or opts to do nothing, after receiving the signal from the sending agent. The lexicon and the set of actions are fixed for the agents, thus the task can be reduced to acquiring the associations between actions and signals. Crumpton implemented two symbol signals but he concluded that such enhancements were not successful.

Another study involving a multi signal scheme belongs to Neubauer (Neubauer, 2004). In his work, the lexicon of interacting agents converge on a common lexicon while describing colors which they perceive through three different color channels. Agents structure categories of colors and relate them to words. Moreover the agents are able to generalize some of the subcategories to a broader category. For instance, if the third channel is commonly below 0.5 for all the colors, subcategories of $0-0.2$ and $0.2-0.5$ are merged to a generalized category of "below 0.5". After common categories and a lexicon to describe those categories emerge, agents use a bag of words, contains three words to describe a color composed of three channels. There is a key conclusion about compositionality in his work. Neubauer (Neubauer, 2004) states that multi word structure emerges only if the environment is structured i.e. the meanings in the environment exhibit a regularity. This is in line with Kirby's findings (Kirby, 2007) in which it is concluded that the compositionality in the experiment only emerges when the meaning space is structured not random. Further details of the study are not given here as it doesn't use language game model.

A key issue in multi word games is the problem of word boundaries. This question has also been studied by language acquisition researchers, in real life. 8 month old infants can correctly identify the beginning and the ending of the words by using the statistical information (Saffran, 1996). Therefore in an experiment it is usually assumed that the parsing of the signals into smaller elements is innate for the agents. Removing this assumption is not a difficult task. Parsing can be done by simply searching the lexicon for possible parses. However there are two drawbacks for this simple method. The first one is possible ambiguities. Secondly, the signal will be perceived as non decomposable, if the lexicon does not contain the sub parts of the whole signal. As a result the agent does not learn the sub parts of the signal (Looveren, 2005). If the experiment includes a complicated syntax, parsing takes place within the model, thus parsing problem is solved.

2.2.2 Meaning space

The meaning space refers to the set of meanings to be expressed. The choice of a meaning space is based on two criteria. Firstly the focus of the experiment is important, e.g. if the case is compositionality, meaning space is tuned between being structured vs unstructured (Kirby, 2007; Laskowski, 2006). Second criterion for the choice of a meaning space is its relation to the real world, i.e. more realistic representations are used dominantly in time. For instance the continuous meaning space is frequently chosed in the designs than discrete meaning space.

One of the common assumptions about the meanings is that the perception of a meaning is innate and universal. It is innate because all the agents in the simulations are assumed to perceive the environment. It is universal because their perception is identical. For instance the color or the size perception is the same among the agents.

Discrete meaning space

The earliest examples of language evolution simulations included discrete meaning spaces. The first example of them, which is also mentioned in the previous section, belongs to Hurford. The meanings, in his case the concepts, are discrete, fixed and shared (Hurford, 1989). Also the studies that focus on syntax and compositionality make use of atomic meanings in order to simplify the semantical side of the issue (Kirby, 2007).

The model of Steels (Steels, 1996) proposes a different meaning representation; a discrete valued vector. As a result the potentially expressed meaning space is still finite but not predefined. The meaning vector contains features with corresponding discrete values.

Continuous meaning space

With a continuous meaning space, the categorization -or discretization, of the meanings is left to agents. This change has two important benefits for the study. It overrides the assumption of discrete meanings, which has already been unrealistic. Secondly, letting the agents to come up with their own internal representation of meanings, such as quantization or categorization, makes agents ore human like (Laskowski, 2006). Besides, the overall result is less predictable as interaction plays a bigger role.

One of the first examples of continuous meaning representation is proposed by Steels (Steels, 1996). In his work, the feature values are continuous. The agents have to divide the meanings into subparts. This division algorithm within each agent is innate and universal for the simulation. This algorithm, which is called discrimination trees, tries to achieve to distinguish objects by using the divisions in the meaning space. Smith (Smith et al., 2003) also used a similar model in which he defines non linguistic communication differently.

Another line of research which makes use of continuous meaning spectrum is prototype based models. Vogt (Vogt, 2000), by using robots, implemented prototypes which are categorical points in the meaning space. In this case, categorization is based on the prototypes, which represent the structured meaning space best. In similar works, the categorization is still discrete. It is discrete because a topic is either a member of a category or not. Moreover, a topic can not be a member of multiple categories. In order to tackle this, Belpaeme (Belpaeme, 2002) proposed another model including graded membership and overlapping categories. His model represents colors only. The recent example of prototype based models belongs to Laskowski, in which he also tried to tackle the discretization problem (Laskowski, 2006).

2.2.3 Topic

In language game models, topic refers to the subject of the language game. If the subject is correctly guessed by the hearer in a language game, the communication ends with success. The topic is an important variant in the simulations.

In the experiments of Steels, topic is a physical object which is defined by a feature vector (Steels, 1996). In Smith's work (Smith et al., 2003) topic is also an object, this time the values of the feature vector are continuous. In Kirby's study on compositionality of language space, the topics are also objects which are defined by short feature vectors filled with discrete values (Kirby, 2007). In the Talking Heads (Steels, 1996) and in the work of Vogt (Vogt, 2000),the objects are physically grounded objects, such as a green square on a whiteboard and the agents are also robots. All of the mentioned experiments contain a finite object set.

In comparison with the experiments that center the emergence of lexicon, the experiments on syntax require more complex topics. In work of the Kirby (Kirby, 1999) the topic is a whole sentence, as one of the issues is the evolution of word order in languages. In the study of emergence of the cases, which is conducted by Steels, the topic is the relation between actions and objects (Steels). In Belpaeme's work, the topic is the color (Belpaeme, 2002).

2.2.4 Context

In language games, context refers to the set of objects which is available for the agents, while communicating. The context can contain either all of the possible topics, e.g. when they are finite, or just a portion of it.

In the works of Steels and Smith, the context contains random, finite, pre-generated objects. In Steels' experiments the context is fixed through out the simulation. In some studies context is not random to analyze the effect of the context. In Kirby's work (Kirby, 2007) context is an important variable. The topics in the context are structured/unstructured and with high/low density. He concluded that compositionality emerges in the structured contexts. In line with Kirby, Webb found that structured context has significant consequences (Webb, 2005). In prototype based models, the context is structured and finite in order to infer clusters in feature/meaning space (Laskowski, 2006).

Last issue to mention about context is the bottleneck effects. The limited context causes a bottleneck effect to form a expressivity and regularity in the language. Although it is not referred as *context*, in ILF context corresponds to the set of meanings. In such studies, the agents are prompted by a set of meanings and they produce a language, based on their hypothesis (Details of ILF is given in Section 2.1.2).

2.2.5 Non linguistic communication

The necessity of non linguistic communication roots from the real life phenomena called as joint attention. In language acquisition of children, joint attention plays an important role. It serves the purpose of agreement on the topic (or object) between the speakers. In child learning, it takes place by pointing or gazing. This ability in children begins to appear very early. Although this ability reduces the uncertainty in real life communication, it can not resolve the uncertainty. There is still ambiguity to some extent, e.g. when the parent points to a dog, does he refer to its ear, legs or color?

In language game models, there are two ways to model this fact. First solution is to model gazing or pointing, as if the referrer refers the whole body of the referent which becomes the topic. This interaction is noiseless, i.e. the hearer correctly interprets the referent. Second option is noisy non linguistic communication. The referrer ambiguously points to the referent which may be one of the topics. Although the former is the most frequent choice of design, the latter is also used in some studies (de Jong and Steels, 2003; Steels and Kaplan, 1998).

2.2.6 Population and social network

The simulation contains agents in an environment. The speaker and the hearer are chosen among the agents. This choice creates a degree of freedom. In many cases, pairing the agents is done randomly. Although in some studies social network is not random and structured (Ke, 2004; Coupe, 2004). In order to resemble real life human groups, clusters are generated. Inter cluster interaction is set to a low level whereas intra cluster interaction is set to a high level. This way, different linguistic properties such as bilingualism (Castello et al., 2008) emerge within different clusters.

2.2.7 Genetic variation

Genetic variation refers to the cognitive abilities of the agents in the next generation. Not all the simulations implement genetic variation. For instance, the simulations in which the population size is fixed, genetic variation does not exist. In Briscoe, the offspring obtains the grammar of her parents, after a series of crossover or mutation (Briscoe, 1997). In MacLannan, the agent shares her parents' signal-action table entries with a crossover (MacLennan, 1992).

2.2.8 Evaluation measures

The termination of a simulation is determined based on the number of the language games or on the simulation's behavior (Laskowski, 2006). However ending simulations either way is not sufficient to conclude the experiment, therefore evaluation techniques are needed to compare different experiments, find out the trajectory and observe the behavior of the evolution in the experiments.

Communicative success

Communicative success is the most primitive measure in language evolution experiments. It shows how well the agents are doing in communicating. In numbers, it is simply the ratio of the successful game count to the number of all games. The timing of the measurement may vary; the measure can be cumulative or instantaneous. In the former, perfect communication score is not be expected, although in the latter the stable languages score hundred percent. Communicative success is necessary in explaining the result of the experiment however it is not sufficient. For example a population may communicate well, but this doesn't indicate anything about the cognitive structure of the agents and the specifics of the language such as lexicon and syntax.

In some studies fitness score is used instead of the communicative success (Steels, 1997). Nevertheless some researchers devised their own fitness score definitions. Perfors defines fitness as the number of correct replacements of the unknowns in the agent's table, in which association of the meanings and signals are kept (Pefors, 2000). The definition of fitness score in Oliphant is similar to the definition of communicative success. Kirby tells in (S., 1996) that Oliphant's definition is more formalized.

Lexical coherence

Lexical coherence measures the agreement between the lexicons of the agents among population. This measurement is calculated by comparing the lexicons of the agents (Looveren, 2005). High values of lexical coherence do not always indicate stability and communicative success; e.g. the agents of the population may utter the same word for every topic yet they do not communicate at all (Looveren, 2005).

Compositionality

The compositionality measure which is proposed by (Smith et al., 2003) is "the degree of correlation between the distance between pairs of meanings and the distance between the corresponding pairs of signals". This correlation is calculated by taking all possible meaning pairs into account $\langle m_i, m_j \rangle$ such that i≠j and their corresponding signals $\langle s_i, s_j \rangle$. Then the distance between those pairs are calculated and the correlation between $\triangle m_{ij}, \triangle s_{ij}$ is computed.

Lexicon size

Lexicon size of agents in different models is compared for issues of compositionality and efficiency. If the lexicon size is small and the model is stable at the end of the simulation, then the agents can express more, relative to the number of topics in the environment (Looveren, 2005). Intuitively it is obvious that relatively low lexicon size is a candidate for high expressivity and compositionality.

Chapter 3

A NEW SIMULATION

This thesis proposes a new language game simulation in this chapter and experiments with the model. The experiments and their results are in the Chapter 4. Having told the related work in this field in Chapter 2, this section first tells the motive behind this new simulation in Section 3.1, and outlines the simulation then specifies the details of it in Section 3.3. Lastly, the chapter concludes with a short discussion of the model in Section 3.4.

3.1 The Motive

This thesis proposes a new model to tackle the following questions.

- Can compositionality emerge without a complex syntax but with a simple bag of words rule, through a language game based simulation?
- Can such a compositionality fill the transition between one word naming to multiple word naming?
- What are the effects and pressures on the emergence of such compositionality?

The methodology to answer these questions is the following.

1. Design a language game, take into account: compositionality and realization
2. Design a language game based model -simulation, in which the language game is played between agents Define measures (metrics) to measure compositionality
3. Run the simulation with the defined parameters
4. Observe if the simulation converges to a communicative population with a compositional language

5. See the effects and pressures on compositionality by changing the parameters

3.2 The story of the simulation

The simulation tries to realize the interacting agents. The agents live in an environment which contains not only the agents but also a set of objects. The agents' only activity is a language game which is as the following. A randomly determined agent sees an object and says some words to describe it. Then another randomly determined agent hears the words and tries to guess the object that the speaking agent talked about. When he decides on an object, he points the object to the speaking agent. If the object is the one that the speaker described, the speaking agent nods to tell that the other agent is right. If the object is not the one that the speaker described, the speaking agent shakes his head to tell that the other agent mistook. The game ends and another game starts somewhere in the environment.

At the end of the simulations, the agents are either good at communicating with others about the objects or they still fail to do so although they did plenty exercise on the game. Besides their success in communication, the size and structure of their vocabulary are evaluated.

3.3 The realization of the simulation

The previous section gave a rough outline of the simulation. This section will discuss the details to implement the story that was told.

3.3.1 The objects and the context

The objects are represented as a vector that contains discrete valued (0 or 1) n features. It is obvious that discrete values are insufficient realizations because the real world is as continuous as our perception. However this simplification in this simulation helps us decrease the complexity and focus on the questions of compositionality and lexicon, by assuming that the representation problem is solved. With such an assumption, the task becomes simpler, although it leaves out the possibility of a combined solution for language evolution and discretization. A different approach, which is not implemented in this study, is using a continuous feature space and incorporating discretization within the learning methods.

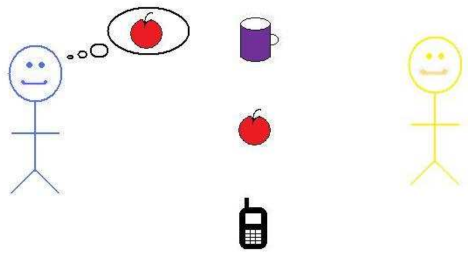

Figure 3.1: An agent picks an object out of all objects

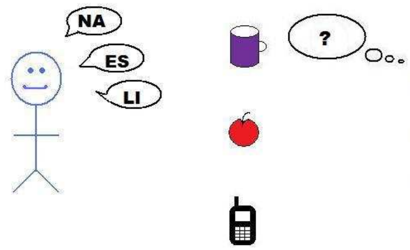

Figure 3.2: The agent says some words to describe the object

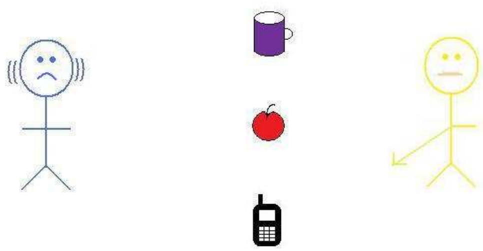

Figure 3.3: Failed game: Another agent hears the words and guesses the object, but the speaking agent shakes his head as the hearer picked the wrong object

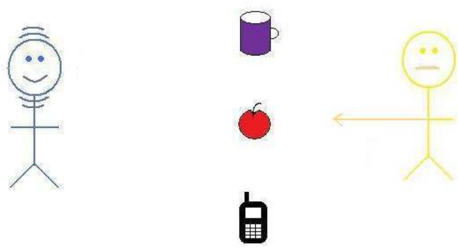

Figure 3.4: Successful game: Another agent hears the words and guesses the object. The speaking agent nods as the hearer picked the correct object

The context is all the objects in the environment. The objects are randomly generated n-tuple vectors. All the objects in the environment are distinct. All the agents can see all the objects at any time.

3.3.2 The words and the word bag

At the beginning the lexicon of the agents is empty. Every agent utters a random bag of words after seeing an object for the first time, in order to initiate a lexicon. The random words which are 5 characters long build their lexicon. Throughout the simulations the meanings of the words evolve. Some of the words may become extinct.

A speaking agent produces a bag of words to describe an object to the hearing agent. The word bag has a length limit. The speaking agent fills the word bag with words one by one. All of the words in the word bag have to be distinct. However the agent may stop putting a new word before reaching the length limit, if the agent does not have any optimal words left in its lexicon. In this case the speaking agent puts the 'null' word into the word bag. The exception of the 'null' word is, the word bag can contain more than one 'null' word, whereas other words can not be repeated. This exception makes the real length of the word bag a variable, which is between 0 and the limit, because 'null' word is not actually a word. At last, the speaking agent says the words in the word bag, without a specific order.

3.3.3 The agents

The agents have two capabilities and a dynamic memory all of which are identical and innate. These traits do not change over time, e.g. size of the memory does not increase or decrease. First capability is describing an object, second one is guessing the described object. The structure of the memory and the capabilities are described in the later sections in detail.

The memory

The memory of the agents contains the history of the games that the agent takes part in. The agent has a limited memory capacity. If the memory reaches that limit, the oldest game in the memory is deleted in order to free space.

The agents remember the following information about the games.

- The object that is used in the game and the object's feature vector
- The word bag that describes the object
- The result of the game (success or failure)

An example of a memory is shown in the Table 3.1.

Game Number	The feature vector	Result of the game	The word-bag
1	10001	0	$W_1W_2W_3$
2	11101	1	W_2W_1null
3	11001	1	$W_7W_6W_4$
4	10001	1	W_1nullW_5

Table 3.1: The memory of an agent after 4 games. The game requires 3 words. The objects have 5 discrete features.

The speaking agent

The available inputs for the speaking agent to describe the object are the object itself, the context and its memory. The desired output is a bag of words. The algorithm proposed for an agent to play the role of speaking agent is given in Table 3.2.

This learning table is given to the decision tree algorithm, specifically to an ID3 (Quinlan, 1986), as the training set. The learning table is built by the contents of the memory and its rows contain the information of object of the game, result of the game and the corresponding word that is told to describe the object. Every game has k words in the word bag, thus it is included in the learning table by k rows. If the memory contains the information of n past games, then the learning table has $n * k$ rows. The input columns of the learning table are the feature of the object and the result of the game. The output column contains the word which described the object in the game. Table 3.3 is a learning table example, which belongs to the agent whose memory is given in Table 3.1. In this example k equals to 3 and n equals to 4, therefore the learning table contains 12 instances to train ID3.

1. Pick an object randomly
2. If the object has never been seen, say a random bag of words and go to 8
3. Build a *learning table* and train a decision tree with the learning table
4. *Classify* the picked object and find a corresponding word by using the trained decision tree
5. Unless the word is the ”null” word, exclude the word temporarily from the memory till the end of the game
6. Unless the word bag is full, go to 3
7. Tell the contents of the word bag without a specific order.
8. See what the hearing agent pointed, and with a non linguistic cue, tell him if he is right or wrong.

Table 3.2: The algorithm of the speaking agent role

In step 4, after training phase, in order to produce a word bag, ID3 classifies an instance. An example of the instance is shown in Table 3.4, where the object, that the agent picked in step 1, has the feature vector of 10101. Besides the feature vector, classified instance contains the field of the game result. The result of the game in the classified instance does not vary, and is always set to 1. The reason behind this choice is that the classified instance needs to be closer to the instances of the successful games. This design choice can be criticized by claiming that the field of game result would be redundant if the learning table contained only successful games. In such a case the learning table would be empty till a game ends successfully. However with the current algorithm, the agent is able to produce an output, even if he never succeeded in a game.

After ID3 gives an output, for instance w_2, the algorithm reaches step 5 If the bag of word is full, the agent says the its contents. If it is not, the algorithm goes back to step 3,

Input Object Features	Input Result of the game	Output Word
10001	0	W_1
10001	0	W_2
10001	0	W_3
11101	1	W_2
11101	1	W_1
11101	1	*null*
11001	1	W_7
11001	1	W_6
11001	1	W_4
10001	1	W_1
10001	1	*null*
10001	1	W_5

Table 3.3: The learning table, the input for the decision tree algorithm

and creates another learning table. The new learning table does not contain the instances whose output is contained by the accumulated word bag. The reason behind this exclusion is preventing the agent to put a word into the bag more than once. The exception of this rule is the 'null' word. The 'null' word is not excluded from the learning table.

In step 7, the speaker has a full bag of words. The speaking agent utters the words in the bag with no specific order. The motive for not keeping the order, by which the words are accumulated, is to break any structural pattern that may be caused by the ordering of the words. As a result, it is ensured that the words in the bag are only connected by the 'and' operator and the order of words does not play any role.

In step 8 the speaking agent tells the hearer that he is right or wrong depending on the object that the hearer pointed. The speaker remembers this actual result of the game.

Input	Input	Output
Object Features	Result of the game	Word
10101	1	?

Table 3.4: The instance to be classified by ID3

The hearing agent

When the agent is the hearer in the game, his task is determining that object that is described by the speaker. The framework of the algorithm for guessing is similar to the framework of speaking algorithm. However the details of the steps differ. The algorithm for the hearing agent is given in Table 3.5.

1. Take the word bag that is told by the speaking agent as input

2. Build a *learning table* and train the decision tree with the learning table

3. *Classify* the word bag, and find the object, to which the word bag refers, by using the decision tree

4. Tell the speaking agent the object with a non linguistic cue

5. Observe the non linguistic expression of the speaker to see if the game is successful or not

Table 3.5: The algorithm of the hearing agent role

The available pieces of information for the hearing agent are its memory and the word bag, which is told by the speaker. In step 2, the hearing agent builds a learning table from its memory. This learning table, which is different from the learning table of the speaking algorithm, contains the previous games' word bags and the results of the games as the input columns. The only output column contains the references to the objects. The number of rows in the learning table equals to the game count of the memory. The Table 3.6 shows an example of such a learning table. The learning table belongs to the agent whose memory is shown in Table 3.1. In this example the length of the word bag is limited to 3 and the objects have 5 features. Trained ID3 by the learning table, finalizes Step 2.

Input	Input	Input	Input	Output
$1^{st}Word$	$2^{nd}Word$	$3^{rd}Word$	Result of the game	Object
W_1	W_2	W_3	0	O_1
W_2	W_1	null	1	O_2
W_7	W_6	W_4	1	O_3
W_1	null	W_5	1	O_4

Table 3.6: The learning table that is built by the hearer

In step 3, ID3 classifies the instance which contains the word bag that is heard and the result of the game as the input fields. The result of the game is set to 1 due the reason explained in Section 3.3.3. Table 3.3.3 is an instance example to be classified by the trained ID3. In this example the word bag, that is heard, contains the words: W_3, W_1 , W_5.

Input	Input	Input	Input	Output
$1_{st}Word$	$2_{nd}Word$	$3_{rd}Word$	Result of the game	Word
W_3	W_1	W_5	1	?

Table 3.7: The instance to be classified by the trained ID3

The classification result is an internal, non linguistic reference to an object. The agent points the object to the speaker. In step 5 he observes the expression of the speaker to see if the game ends with a communicative success or not. The agent remembers the game with this actual result.

3.3.4 Parameters, evaluation and experiments

The experiments are designed to observe the effects of the parameters on the evaluation measures. The parameters are as the following.

- The length of the word bag
- The number of objects in the environment
- The number of objects in the context
- The length of the feature vector
- The agent's memory capacity
- The population size

Some of these parameters are dependent, e.g. the object count depends on the feature count, because the latter is limited by the former. Bearing these dependencies in mind, the parameters will vary in different runs of the simulations. In the experiments, the following evaluations will be done.

- The communicative success
- The compositionality measure proposed in (Kirby, 2002)
- The compositionality measure proposed in this dissertation (it will be referred as the second measure of compositionality)

Communicative success is the average communicating score between all agent pairs. The communicating score between a pair of agents is retrieved over all objects in the context. As mentioned in part 2.2.8, there are two different ways of measuring the communicative success. First one is cumulative, which computes the overall score for the whole simulation and second one is instantaneous, which computes the communicative success of the population at a given time. By using the first way of measuring, the value can almost never be calculated as 1.0, due to the failures in the games at the early stages. However it converges to 1.0, in a perfectly communicating population. On the other hand, instantaneous measure is better to analyze the trajectory of the simulations, because the measured values are comparable through the simulation.

The compositionality of a language can be measured by the method in (Kirby, 2002), which is described in the Section 2.2.8. However this measure has to be modified for this simulation. Because the original definition includes the distance between two signals, whereas in this study there are word bags instead of signals. As this definition is incompatible with the word bags, the distance between two word bags needs to be defined. For this simulation the following distance measure is proposed. Suppose that s_i and s_j are a pair of word bags and contain the words $\langle w_{i1}, w_{i2}, w_{i3} \rangle$ and $\langle w_{j1}, w_{j2}, w_{j3} \rangle$ respectively. The distance between the word bags are $\sqrt{length(wordbag) - size(contained_{ij})}$, where $contained_{ij}$ is the intersection set of s_i and s_j. The other definitions in the Section 2.2.8 are kept unchanged. Lastly the measured compositionality values of all the agents are averaged.

Second compositionality measure that is proposed by this thesis is: $\frac{compositional-word-count}{lexicon-size}$. The compositional word count is the number of distinct words that are contained by all s_i and s_j word bag pairs, where the word bags are told by the agent for two different objects i and j.

3.4 Discussion of the model

There are two issues which need to be discussed. First issue is the compatibility of the model with the cognitive aspects. This issue is observable in two points, first of them is the learning schemes and the second point is the guessing algorithm. In the model, learning schemes of the speaker and the hearer include different algorithms. The algorithms differ not only by their functionalities, which are inherently different, but also by their details in learning. For instance, their learning tables are totally different. On the other hand it may be argued that humans do both tasks of speaking and guessing, in a similar manner and use the same learning scheme. Second point is that hearing agent guesses without knowing explicitly what words mean, whereas a human would determine the described object by assessing the meanings of the words. Although there are some intuitions behind the argument of incompatibility, this model is a possible explanation of the emergence of a compositional lexicon, because how human mind solves the problem is still an open question.

Second issue is the 'null' word. The reason behind such a design decision is releasing the pressure on the speaking agent to speak, in case that there are no optimal words left. Moreover the exception of the 'null' word creates a degree of freedom in the length of the word bag (see Section 3.3.2 for details.) Although the iteration of putting a word into the word bag is done until the length limit is reached, the count of non null words may not be equal to the length limit. As a result, the agents may end up using a varying length of word bags. The length of a word bag becomes a variable to study.

Chapter 4

EXPERIMENTS AND RESULTS

This chapter contains the information on 4 experiments with model presented in Chapter 3. First of these experiments demonstrates the base result of this thesis. Following experiments modify some of the parameters to see their effect on the evaluation measures.

4.1 Experiment 1: The new model leads to a stable language with high communicative success and compositionality

The first experiment is conveyed to test the new model with a set of default options for stability. The parameters can be seen in Table 4.1.

A run of the experiment starts with an initial state and ends when the communicative score exceeds 91.0%. To evaluate the experiment, after every 500 games, the communicative success and the compositionality are measured. In order to decrease the effect of chance, this run is repeated for 10 times. The instantaneous measured values of each run are averaged. The evaluations are given in sections 4.1.1 and 4.1.2. Besides the quantitative analysis in these sections, Section 4.1.3 contains a qualitative analysis of the ID3 representations to depict the evolution of meanings.

Parameter	Value
Agent Count	10
Object Count	10
Length of the feature vector	4
Length of the word bag	5
Memory size	200

Table 4.1: The parameters of the first experiment

4.1.1 The communicative success

It is observed that, at the end of each run, the agents can communicate about every object in the context. The communicative success converges to 1.0 as the agents play more games. The result can be seen in Figure 4.1

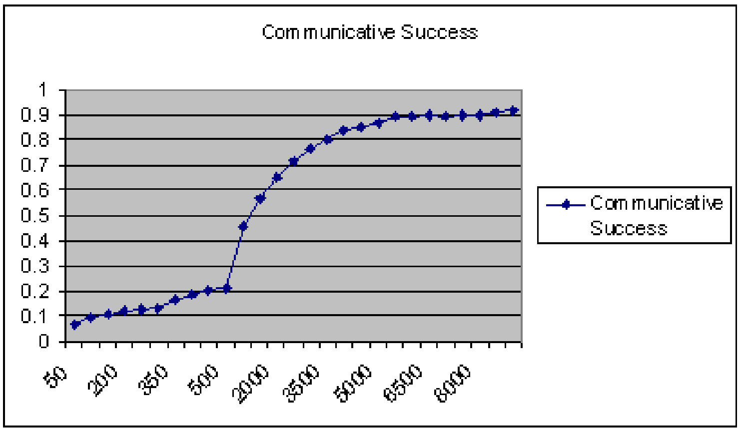

Figure 4.1: Measured communicative success, after every 500 games (Standard deviation of the error < 0.001)

4.1.2 The compositionality

The computed values by using the compositionality measure of Kirby, which is explained in 2.2.8, show that a high compositionality exists even after the first iterations of the game, as it is seen in Figure 4.2.

If we compare the compositionality of Kirby with the baseline, which is nearly 0 correlation, it is clear that compositionality is high above baseline.

If the second measure of compositionality is used, which is described in 3.3.4, it is clear that 20 % of the words are used to describe other objects, which signals compositionality. The result can be seen in Figure 4.4.

4.1.3 The Decision tree, ID3

This section makes a qualitative analysis on compositionality, by depicting an agent's internal meaning representation, which is an ID3. The Figure 4.5 shows the state of the ID3,

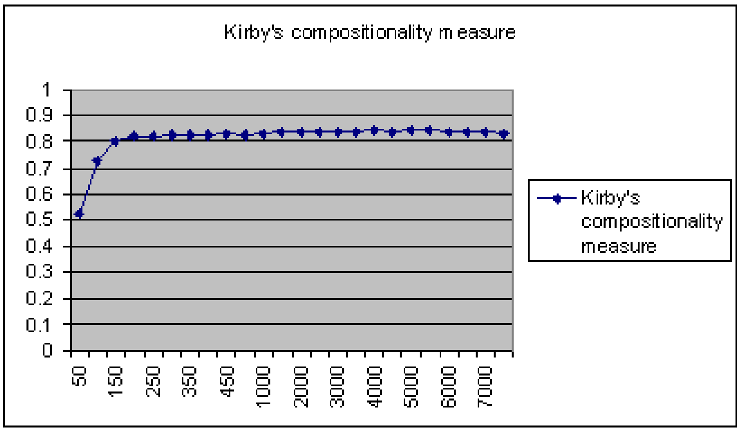

Figure 4.2: Kirby's compositionality measure after each 500 games (Standard deviation of the error < 0.001)

after 7000 games when the communicative success of the population is 91 % . The size of the observed agent's lexicon is 6. This shows that, for this specific experiment, 6 words are sufficient to describe 4 featured, 10 objects. This is not perfect compositionality, although it signals compositionality significantly.

4.2 Experiment 2: Compositionality peaks when the number of words equals to the number of features

Section 4.1 shows that the model leads to compositionality and stability for a given set of parameters. This experiment tunes one of the parameters, the length of the word bag, in order to see its effect on compositionality. Fixed parameters are shown in Table 4.2.

Parameter	Value
Agent Count	10
Object Count	10
Length of the word bag	5
Memory size	200

Table 4.2: The parameters of the second experiment

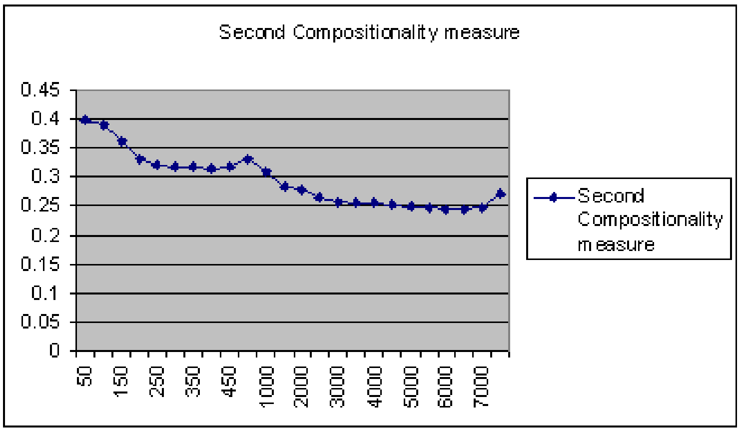

Figure 4.3: Second compositionality measure, computed after each 500 games (Standard deviation of the error < 0.001)

The length of the word bag is the variable of this experiment. The length of word bag is incremented by 1 in each run, starting from 1 to 7. All 8 iterations continue till the agents are communicatively successful.

In this experiment, the number of games needed for a stable language and both compositionality measures are computed.

The main result of this experiment is that the compositionality peaks when the length of the word bag equals to the number of features. The values for Kirby's compositionality measure are analyzed in Figure 4.6. It is seen that it peaks when the length of the word bag is 5, which is equal to the number of features. The evidence from the second compositionality measure is also in line with the result of the first compositionality measure. The graph in Figure 4.8 contains the values for the second measure. The peak of the graph is when the length of the word bag is 5.

Another observation about compositionality is the emergence of a holistic language when the length of the word bag is 1. When the word bag contains only one word, the language obviously becomes holistic i.e. all the objects are associated with a single word which can not be decomposed. The first evidence that supports this hypothesis is from the computed values of Kirby's compositionality measure. The graph in Figure 4.6 shows the varying values for compositionality which is at its lowest level when the length of word bag is 1. Secondly, the graph in Figure 4.8 also shows that, the compositionality is 0, for the first

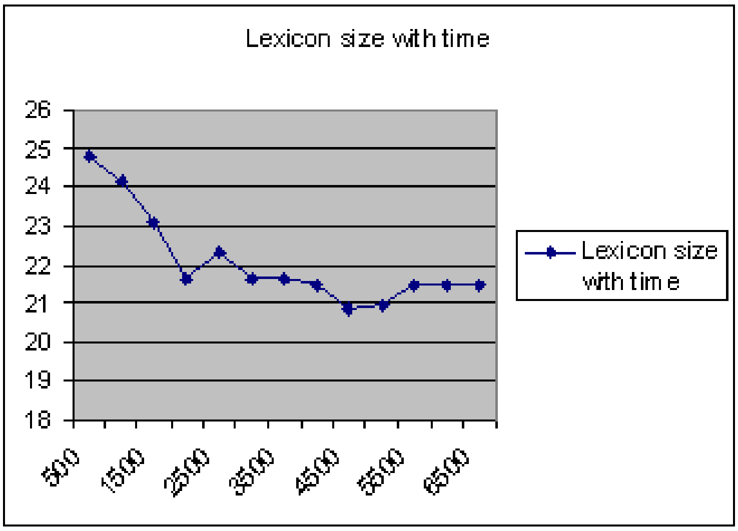

Figure 4.4: Lexicon Size with time

value of the length of the word bag. This is an expected result when the language is completely holistic. Last evidence is a qualitative one. The ID3 in Figure 4.7 belongs to an agent at an instance when the communicative success is 95 % and the word bag length is limited to 1. It is seen that the size of the lexicon equals to the number of objects, which is 10. To sum up, the stable language is holistic, when the length of the word bag is 1.

This experiment is an opportunity to examine and compare two compositionality measures. First observation is that Kirby's compositionality measure does not get below 0.63 which is significantly high for a correlation measure. This shows that even holistic languages score high with Kirby's compositionality. Thus the values of the measure may not be significant whereas they are useful for comparison. On the other hand, the second compositionality measure gives better results. For instance the measured value is 0, when the language is on the farthest point from compositionality i.e. the length of the word bag is 1. Thus it is a good candidate for being a better compositionality measure rather than the measure proposed in (Kirby, 2002).

Last conclusion of the second experiment is derived from the amount of games needed to reach a stable language. Previously, it is shown that the languages, which have a shorter word bag in the language game, are merely compositional. On the other hand, less compositional languages require fewer games to reach stability in comparison with more com-

```
feat0 = true
|   feat3 = true
|   |   feat2 = true: VVGIV
|   |   feat2 = false
|   |   |   feat1 = true: UYIHJ
|   |   |   feat1 = false: CLSOA
|   feat3 = false
|   |   feat2 = true
|   |   |   feat1 = true: null
|   |   |   feat1 = false: CLSOA
|   |   feat2 = false: PDIUM
feat0 = false
|   feat1 = true
|   |   feat2 = true: OAFFX
|   |   feat2 = false: AVHJV
|   feat1 = false: OAFFX
```

Figure 4.5: State of the ID3 of an agent, when the language is stable, after 7000 games

positional ones. The graph in Figure 4.9 supports this conclusion. This conclusion brings a trade off between compositionality and effort. Although this thesis does not address this issue, at least for this simulation, there has to be another pressure on the language to be more compositional because holistic language requires less effort.

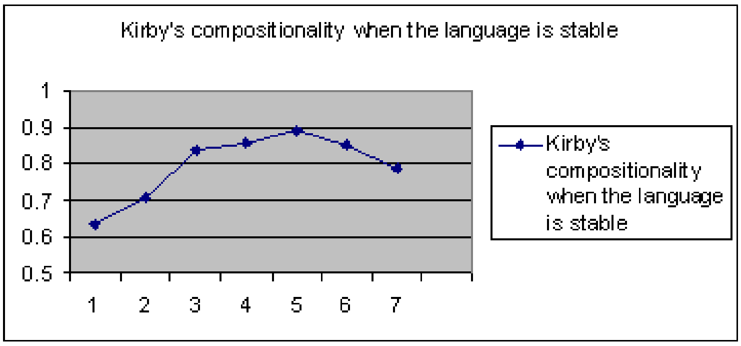

Figure 4.6: Kirby's compositionality based on the word bag size

```
feat2 = true
|   feat3 = true
|   |   feat0 = true: QQDMT
|   |   feat0 = false: null
|   feat3 = false
|   |   feat1 = true
|   |   |   feat0 = true: HAJSK
|   |   |   feat0 = false: JFKAX
|   |   feat1 = false: DSDHR
feat2 = false
|   feat4 = true
|   |   feat1 = true: ELOIA
|   |   feat1 = false
|   |   |   feat3 = true: MUMWC
|   |   |   feat3 = false: DJSON
|   feat4 = false
|   |   feat1 = true: PTFYH
|   |   feat1 = false: FJIIK
```

Figure 4.7: The state of ID3 of an agent, when the language is stable and word bag length is 1

4.3 Experiment 3: Fewer objects in context creates bottleneck effect on compositionality

This experiment aims to find out the effect of the number of objects in the context on compositionality. This effect is observed by running experiments for different numbers of context objects. The other parameters seen in Table 4.3 are fixed. The variable increments by 2 from 4 to 16. For each value of the variable, the experiment is run 10 times till the language reaches stability. After termination, the compositionality measures are computed and they are averaged to reduce the effect of chance.

The main result of this experiment is that if the context contains fewer objects, the agents develop a more compositional lexicon according to both compositionality measures. Trajectory of compositionality for different values of object count is seen in Figure 4.10. Compositionality is calculated by using both formulas and they are indicated in the same figure. This conclusion is in line with the previous research on the bottleneck effect. Differently from the previous work, the effect on the evolution of compositional lexicon is now shown.

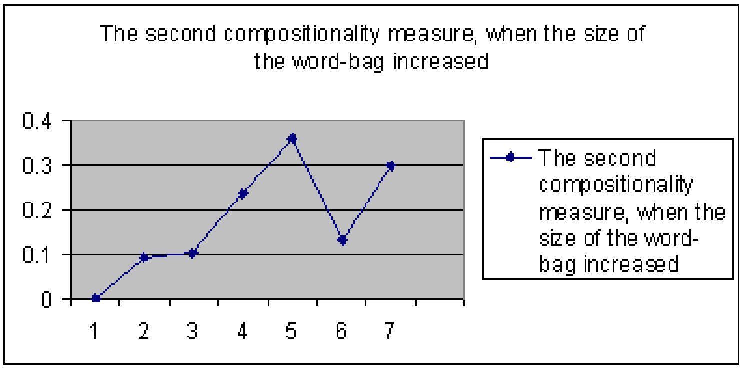

Figure 4.8: Second compositionality based on the word bag size

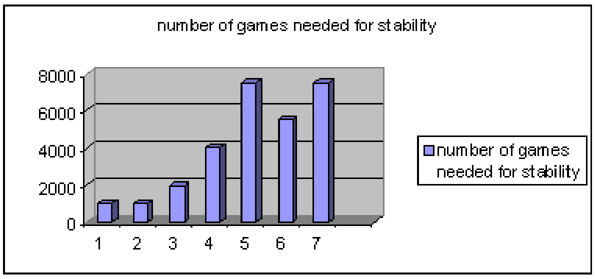

Figure 4.9: The number of games needed for a stable language, variable is the size of the word bag

4.4 Experiment 4: In the model, the agents reach a reasonable agreement on the descriptions of the off context objects

This experiment tests the model if the agents produce similar word bags for the non context objects. The parameters of the experiment are given in Table 4.4. All the parameters other than the number of objects in the context are fixed. The number of the context objects is the variable to observe the effect of the bottleneck on developing a language that covers all possible objects. The number of features is 4, thus the number of possible distinct objects is 2^4. The varying number of objects in the context goes from 4 to 14 increased by 2.

The previous experiment shows that, in the runs of simulations, the agents can not only manage to reach a high communicative success on the context objects but also build a compositional lexicon. On the other hand, this result does not inform about their communicative success if they were tested on the off context objects. This experiment aims to

Parameter	Value
Agent Count	10
Number of features	4
Size of the word bag	4
Memory size	200

Table 4.3: The parameters of the third experiment

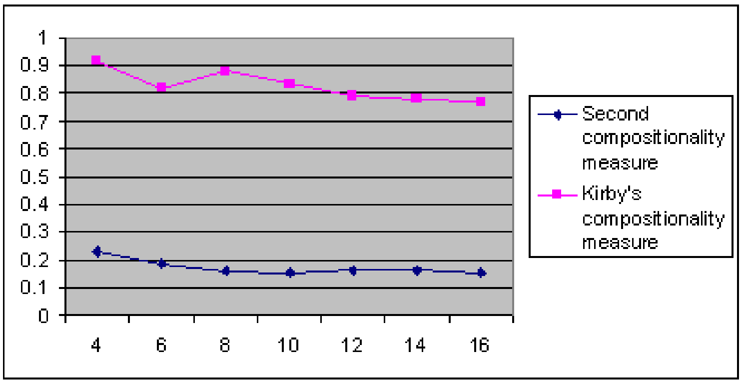

Figure 4.10: The compositionality measure for varying the size of the context

analyze their agreement on the word bags that are created for off context objects, after the language reaches stability.

In order to test the agents on off context objects, a method needs to be devised. This method should measure how much the agents are in agreement on the descriptions of the objects. The proposed method measures the agreement between each agent pair and averages the values. The agreement between two agents is computed as the following. Let i and j be two agents and s_{io}, s_{jo} be a pair of word bags retrieved by asking the agents the description of all off context objects each of which is denoted by o. The agreement between a pair of agents becomes the average similarity of the word bag pairs, s_{io}, s_{jo}. The similarity of two word bags is the ratio of the size of the intersection set to the length of the word bag.

For comparison purposes, the same agreement measure is used with context objects. Besides, the communicative success of the population is computed.

The chart in Figure 4.11 shows the partial agreement on the descriptions of both context

Parameter	Value
Agent Count	10
Size of the word bag	4
Length of the feature vector	4
Memory size	200

Table 4.4: The parameters of the fourth experiment

and off context objects. The context contains 10 objects and the off context contains 6 objects. X-axis is the total number of games played.

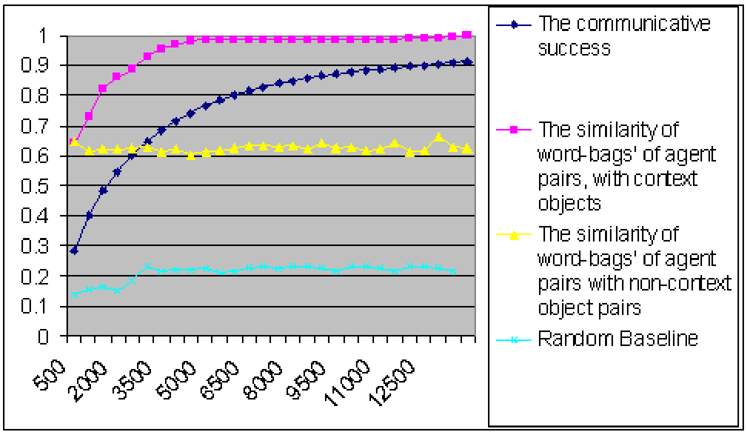

Figure 4.11: The agreement on the descriptions of both context and off context objects. The context contains 10 objects. (Standard deviation of the error < 0.001)

It is seen in the chart that the agents' agreement on context objects is rapidly increases. As a consequence the communicative success among the agents also inclines. At first sight, the slight improvement in average agreement on off context objects may not seem to be a promising result. However those objects were never included in the games that the agents played for many times. Therefore it shouldn't be surprising that the measure on off context objects keeps its line at the same level. On the other hand, the level of the line is far from a random baseline. This means after the language becomes stable, the agents are able to not only communicate about the context objects perfectly but also communicate about the off context objects to random baseline. Random baseline computed with random word bags out of the lexicon of the agents.

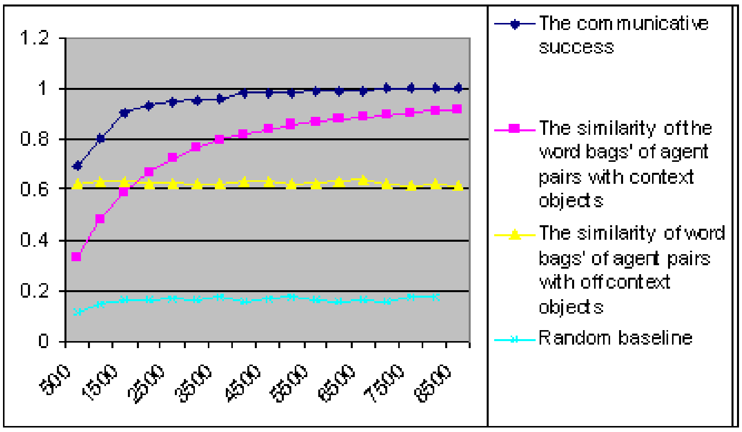

Figure 4.12: The consensus on the descriptions of both context and off context objects. The context contains 6 objects. (Standard deviation of the error < 0.001)

Chapter 5

CONCLUSIONS

This section discusses and evaluates the study in the following outline. The section 5.1 discusses the work in comparison with the previous studies. The section 5.2 summarizes the contribution of this study, with the obtained results. The section 5.3 evaluates the study and model if the work reaches an expected level for a language evolution simulation. The section 5.4 gives a basic plan for the future work.

5.1 Comparison with the previous studies

The comparison with previous studies is not straightforward, because some aspects and parts of this thesis do not have counterparts in other studies. However the results are comparable to some extent. The comparison made with the previous base results of the experiment, research of compositionality and other MNG simulations.

Firstly, in line with the past research, this study shows that after an iteration period, a population of agents reaches to a significant degree of communication when the initial states of the agents are empty. The previous work also reached this result. Nonetheless final result does not say much about the trajectory of this convergence on its own. For example the amount of games needed for stability differs from previous research. As the Section 4 explains, the amount of games needed depends on different parameters. The parameters that affect the rate of the convergence, behave similarly as they did in the works of (Steels, 1997, 1996; Looveren, 2005). When the parameters, such as the number of agents, objects and features increase the convergence time also increases.

If this work is compared with the previous work on compositionality, there are important differences. First of all, many compositionality research incorporates a complex syntax (Kirby, 1999, 2000). On the other hand, there are a few studies that focus on compositionality without syntax similar to this thesis. If those are compared with this thesis, there are some similarities. The work of Batali, resembles this thesis based on the approach

(Batali, 1998a). Batali also worked on the association of features with tokens. But his work was pre-wired e.g. the meaning set was predetermined and the length of the signal was fixed. The meanings in this thesis are dynamic, and are based on the objects in the context. There are also differences in methodology, where Batali's work (Batali, 1998b) is using trained neural networks and ILF, rather than language game simulations. Kirby's work on compositionality of meaning spaces (Kirby, 2007) is also non comparable in some aspects. His work was inter generational, whereas this thesis focuses on intra generational evolution of language. Secondly in (Kirby, 2007) the existence of compositional and holistic languages are assumed before hand to analyze their advantages in the ILF.

Being one of the rare MNG studies, the main contribution of this model to language evolution is on the field of compositionality without a complex syntax. Such MNG simulations previously showed that compositionality may emerge without a complex syntax. The only syntactical rule in this study is a word bag, which is sometimes referred as the 'and' operator (Looveren, 2005). This study, which is line with previous studies, proposes a more general framework, because in previous MNG simulations the focus is very limited such as color (Belpaeme, 2002; Neubauer, 2004).

5.2 Contribution

There are three main contributions of this dissertation. First of them is its explanation for the transition of the language from single words to multiple word chunks. Second contribution is the results of the experiments on compositional lexicon emergence. Last one is its methodology.

First and main contribution is that this work offers a transition scenario from single word languages to multiple word languages. Previous studies proposed plausible explanations for the emergence of lexicon. In these studies there were no explanations about how the words get together to express a wider set of meanings. That is why; many studies on the emergence of syntax have been carried out. The Section 2 contains the detailed literature review. This work is between syntax and lexicon studies and it offers another account of explanation which is as the following. This work shows that, with the model proposed here, the words can combine to form different meanings. Although the model does not explicitly enforce a rule for combining words, a rule emerges from the learning schemes. This rule is a simple 'and' predicate. The 'and' predicate is the proposed explanation for the transition from single word naming to a syntactical language, although it signals a very simplistic syntax which contains only one rule.

Second class of contributions consists of the results of the experiments. The first experiment demonstrates the base result which is discussed in the previous paragraph. Second experiment shows that compositionality in the lexicon peaks when the length of the word bag equals to the feature count. This information theoretical result is significant because it tells that the agents tend to use compositional words more, if the number of words in their word bag equals to the feature count. Third experiment observes the known bottleneck effects, this time they are depicted for compositionality of the lexicon. The conclusion is that if the agents are shown fewer objects, they tend to have a more compositional lexicon. Last experiment shows that the agents in the model do not only agree perfectly on the word bags of the context objects but also agree to a significant extent on word bags of the off context objects. This conclusion is important because the model leads to agents that have similar inner representations for meanings.

Last remark about the results has to be made on the cost of compositionality. In the first experiment it is shown that the compositionality comes with a price of more language games i.e. more exercise and time. Hypothetically speaking, if there was a pressure of time and effort on population, compositional language would not emerge. This result is significant, because, for at least this model, there has to be other tendencies to lead the language among the agents towards building a compositional lexicon.

In the methodology of the thesis, there are a few new approaches used, although most aspects of the work resemble previous language game simulations. First of them is using a machine learning algorithm in the learning schemes of the agents. Second one is the unified non linguistic and linguistic cognitive aspects of learning.

Although there is plenty of previous work which uses machine learning algorithms such as neural networks, in the language game models there are not any specific type of algorithms. At first sight, it may be claimed that replacing these specific algorithms with a ML algorithm is not a novel approach. This criticism is right to an extent, because based on the results; the functionality of the algorithm is similar to the previous ones. However using an algorithm which is well described, conventional, multi purpose and which roots in ML background has some other benefits. First of all, when the learning framework is transformed to a ML background, all the tools in this literature such as discretization, quantization, filtering etc. are available. Secondly when the framework of the learning is standardized, it is easy to replace the algorithm with another ML algorithm in order to see if the other algorithm can produce similar or better results. Lastly, using a conventional ML algorithm rather than a specially tailored algorithm is simpler and more convenient.

Last contribution about the methodology is interesting in terms of the debate on nativism vs non nativism. In language game models, the learning schemes of the agents contain algorithms of which discrimination game (Steels, 1996; Looveren, 2005; Vogt, 2000) and prototypes (Laskowski, 2006) are the examples of. These examples are designed to replace a non linguistic skill. This skill is distinguishing the objects in the environment with their discriminating features or categorizing the objects, as in the case of prototype models. It is obvious that these algorithms are not language related. An important observation about these algorithms are they the models that root in non nativist camp, because the cognitive skill used in the language game is non linguistic. Linguistic side of the learning scheme takes place after the non linguistic algorithm finds the optimal solution e.g. discriminating all the objects. In other words, the words in the language emerge after meaning representations, which are produced by non linguistic processes, emerge.

On the other hand the model in the thesis takes a different line, by using an ID3 as the learning scheme. ID3 tries to find the features that distinguish objects, as the discrimination algorithm does. Its difference from the discrimination algorithm is that ID3 co-evolves with the lexicon. The meanings of the words change throughout the simulation. The framework in which ID3 is used can be claimed neither as non linguistic nor linguistic whereas the discrimination tree is non linguistic. Therefore this model is closer to the nativist camp.

5.3 Evaluation

A language evolution simulation is evaluated on the basis of its compatibility with the reality as well as its success in producing a stable result. This thesis achieves the latter aim. However the success in the former aim needs to be questioned.

First of all, the agents and the structure of the game are close to reality as it is proven by their survival through plenty of criticism after being implemented in the previous studies.

On the other hand, the objects and the algorithm that the agents use to do the task may not seem realistic at first sight.

The assumption of discrete features has been attacked by some studies (Steels, 1996). To some extent this criticism is right. On the other hand, this assumption is a reasonable choice in order to simplify the model by which focusing on other issues such as agent capabilities is possible.

The learning schemes of the agents cause the biggest question mark about the compatibility of the model with the reality. Firstly, the hearing agent only knows the association of the word to an object reference, without knowing the explicit meaning of that word. On the other hand, humans seem to have the knowledge of the explicit meanings of the word. In other words, they can assess the meaning of a word without creating any association to an object, unlike the hearing agent. Secondly, totally different algorithms of the hearer and the speaker can be seen as unrealistic, when compared with humans. But it should be noted that, to do well in the language game, humans may be using different algorithms, too. As it is not absolutely known how human mind solves the problem this is a possible hypothesis.

5.4 Future work

The future work can move in two important directions. Firstly, a continuous feature space can be incorporated to the model, instead of a discrete one. In order to achieve this goal without major changes in the model, discretization algorithms can be a solution for this transition. However it should be noted that these algorithms are not always successful in machine learning problems. Therefore it should be tested to see whether such a future work produces similar results or not.

Second direction is transforming the cognitive skills of the agents to a more realistic level in order to overcome the problems that are mentioned in the Section 5.3. Similarity between the agent approach and the human approach to a task, will enlighten more on the evolution of a natural language.

BIBLIOGRAPHY

J. Batali. Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy, and Knight C., editors, *Approaches to the Evolution of Language: Social and Cognitive Bases*, pages 405–426. Cambridge University Press, Cambridge, 1998a.

J. Batali. The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In *Beyond Grammar: an experience-based theory of language.* CSU Publications, 1998b.

T. Belpaeme. *Factors influencing the origins of colour categories.* PhD thesis, Vrije Universiteit Brussel, 2002.

D. Bickerton. Catastrophic evolution: The case for a single step from protolanguage to full human language. In J. R. Hurford, M. Studdert-Kennedy, and Knight C., editors, *Approaches to the Evolution of Language: Social and Cognitive Bases.* Cambridge University Press, Cambridge, 1998. URL `http://www.isrl.uiuc.edu/ amag/langev/paper/bickerton98catastrophicEvolution.html`.

E. J. Briscoe. Co-evolution of language and of the language acquisition device. In *Proc. of 35th Assoc. for Comp. Ling.* Morgan Kaufmann, 1997. URL `http://www.isrl.uiuc.edu/ amag/langev/paper/briscoe97acl.html`.

E. J. Briscoe. Grammatical acquisition and linguistic selection. In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, chapter 9. Cambridge University Press, 2002. URL `http://www.isrl.uiuc.edu/ amag/langev/paper/briscoe02grammaticalAcquisition.html`.

A. Cangelosi and D. Parisi. The emergence of language in an evolving population of neural networks. In *In Proceedings of the 18th Conference of the Cognitive Science Society*, 1996.

X. Castello, V. M. EGUILUZ, M. S. Miguel, J. Loureiro-Porto, R. Toivenen, J Saramaki, and K. Kaski. Modelling language change competition: bilingualism and complex social networks. In *Proceedings of 7th Evolution of Language Conference*, 2008.

N. Chomsky. *Aspects of the theory of syntax*. MIT Press, 1965.

B. Comrie and T. Kuteva. The evolution of language and elaborateness of grammar: the case of relative clauses in creole languages. In *International Conferences on the Evolution of Language*, 2004.

C. Coupe. A unified computer model for internal and external constraints in language evolution. In *Proceedings of 5th Evolution of Language Conference*, 2004.

J. J. Crumpton. Evolution of two symbol signals by simulated organisms. Master's thesis, December 1994. URL `http://www.isrl.uiuc.edu/ amag/langev/paper/crumpton94evolutionOf.html`. This thesis reports experiments on the factors promoting or inhibiting the evolution of simulated organisms using strings of length 2 for communication. The simulation program is also available.

I. Davidson. *Language Evolution*, volume 100, chapter The archeological Evidence of Language Origins: State of Art, pages 788–791. 2003.

B. de Boer. Generating vowel systems in a population of agents. In P. Husbands and I. Harvey, editors, *ECAL97*, Cambridge, MA, 1997. MIT Press.

B. de Boer. *The Origins of Vowel Systems*. Oxford University Press, 2001. URL `http://www.isrl.uiuc.edu/ amag/langev/paper/deboer01theOrigins.html`.

B de Boer. Evolving sound systems. In Angelo Cangelosi and Domenico Parisi, editors, *Simulating the Evolution of Language*, chapter 4, pages 79–97. Springer Verlag, London, 2002. URL `http://www.isrl.uiuc.edu/ amag/langev/paper/deboer01evolvingSound.html`.

E. D. de Jong. *Autonomous formation of concepts and communication*. PhD thesis, Vrije Universiteit Brussel, 2000.

E. D. de Jong and L. Steels. A distributed learning algorithm for communication development. *Complex Systems*, 14:315–334, 2003.

I. C. R. Ferrer and R. V. Sole. Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences of the United States of America*, 100(3): 788–791, 2003.

R.A. Gardner and B.T. Gardner. Teaching sign language to a chimpanzee. *Science*, 165: 664–672, 1969.

L. Heidi. Artifical symbol systems in dolphins and apes: analagous communicative evolution or evidence for basic communicative rules? In *Proceedings of 7th Evolution of Language Conference*, 2008.

J. Hurford. Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua*, 77(2):187–222, 1989. doi: 10.1016/0024-3841(89)90015-6. URL `http://www.isrl.uiuc.edu/ amag/langev/paper/hurford89biologicalEvolution.html`.

E. Hutchins and B. Hazlehurst. How to invent a lexicon: The development of shared symbols in interaction. In N. Gilbert and R. Conte, editors, *Artificial Societies: The Computer Simulation of Social Life*, pages 157–189. UCL Press: London, 1995. URL `citeseer.ist.psu.edu/hutchins95how.html`.

J. Ke. Language change and social networks. In *Proceedings of 5th Evolution of Language Conference*, 2004.

S. Kirby. *Function, Selection, and Innateness*. Oxford, 1999.

S. Kirby. Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, editor, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pages 303–323. Cambridge University Press, 2000. URL `http://www.isrl.uiuc.edu/ amag/langev/paper/kirby00syntaxWithout.html`.

S. Kirby. Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation*, 5 (2):102–110, 2001.

S. Kirby. Learning, bottlenecks and the evolution of recursive syntax. In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and*

Computational Models, chapter 6. Cambridge University Press, 2002. URL `http://www.isrl.uiuc.edu/ amag/langev/paper/kirby02learningBottlenecks.html`.

S. Kirby. *The evolution of meaning-space structure through iterated learning*, chapter 253 - 268. Springer Verlag., 2007.

S. Kirby and H. Brighton. Understanding linguistic evolution by visualizing the emergence of topographic mappings. *Artifical Life*, 2006.

C. Laskowski. Prototype categorisation and the emergence of a lexicon in an infiniteworld. Master's thesis, University of Edinburgh, 2006.

J. V. Looveren. *Design and Performance of Pre-Grammatical Language Games*. PhD thesis, Vrije Universiteit Brussel, 2005.

B. MacLennan. Synthetic ethology: An approach to the study of communication. In Christopher G. Langton, Charles Taylor, J. Doyne Farmer, and Steen Rasmussen, editors, *Artificial Life II*, pages 631–658. Addison-Wesley, Redwood City, CA, 1992. URL `citeseer.ist.psu.edu/maclennan91synthetic.html`.

L. B. P. MacNeilage and M. Studdert-Kennedy. Selforganizing processes and the explanation of phonological universals. In B. Comrie Butterworth, G. and O. Dahl, editors, *Explanations for Language Universals*, pages 181–203. Berlin, 1984.

J. Maynard Smith and E. Szathmry. *The Origins of Life: From the Birth of Life to the Origin of Language*. Oxford University Press, 2000.

N. Neubauer. Emergence in a multiagent simulation of communicative behaviour. Master's thesis, University of Osnabrck, December 2004.

M. N. Nowak, J. B. Plotkin, and Jansen V. A. A. The evolution of syntactic communication. *Nature*, 2000. URL `/ref/kirby/nowak00.pdf`.

A. A. Pack and L. M. Herman. Bottlenosed dolphins (tursiops truncatus) comprehend the referent of both static and dynamic human gazing and pointing in an object-choice task. *Journal of comparative psychology*, 2004.

A. Pefors. Simulated evolution of communication: the emergence of meaning. Master's thesis, Stanford University, 2000.

R. Quinlan. Induction of decision trees. *Machine Learning*, 1(2):81–106, 1986.

Kirby S. Natural language from artificial life. *Journal of Memory and Language*, 35:606–621, 1996.

Newport E.L. Aslin R.N. Saffran, J.R. Statistical learning by 8-month old infants. *Science*, 274:1926–1928, 1996.

Rumbaugh D.M. McDonald K. Savage-Rumbaugh, E.S. Language learning in two species of apes. *Neuroscience and Biobehavioral Reviews*, 9:653–665, 1985.

K. Smith. The cultural evolution of communication in a population of neural networks. *Connection Science*, 14(1):65–84, 2002. doi: 10.1080/09540090210164306.

K. Smith, S. Kirby, and H. Brighton. Iterated learning: a framework for the emergence of language. *Artificial Life*, 9(4):371–386, 2003.

L. Steels. Simulating the evolution of a grammar for case luc steels.

L. Steels. Proceedings of the simulation of adaptive behaviour conference. The MIT Press, 1996.

L. Steels. The synthetic modeling of language origins. *Evolution of Communication*, 1(1):1–34, 1997. URL `http://www.isrl.uiuc.edu/ amag/langev/paper/steels97theSynthetic.html`.

L. steels and F. Kaplan. Bootstrapping grounded word semantics, 2002. URL `citeseer.ist.psu.edu/steels99bootstrapping.html`.

L. Steels and F. Kaplan. Stochasticity as a source of innovation in language games. In *Proceedings of the Conference on Artificial Life VI (Alife VI)*, Los Angeles, California, 1998.

M. Tomasello. The social bases of language acquisition. *Social Development*, 1(1):67–87, 1985.

W. J. Turkel. The learning guided evolution of natural language. In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, chapter 8. Cambridge University Press, 2002. URL `http://www.isrl.uiuc.edu/ amag/langev/paper/turkel02theLearning.html`.

P. Vogt. *Lexicon grounding on mobile robots.* PhD thesis, Vrije Universiteit Brussel, 2000.

A. Webb. Modelling type-denoting concepts and words in a simulation of vocabulary development. Master's thesis, University of Otago, 2005.

H Yamauchi. The difficulty of the baldwinian account of linguistic innateness. In J. Kelemen and P. Sosk, editors, *ECAL01*, Lectures Notes in Computer Science, pages 391–400, Prague, September 10-14 2001. Springer.

H. Yanco and L. Stein. An adaptive communication protocol for cooperating mobile robots. In *From Animals to Animats 2: Proceedings of the Second International Conference on Simulation of Adaptive Behavior*, 1993.

VITA

AHMET ENGIN URAL is born in Izmir, Turkey on April 1, 1983. After graduating from Bornova Anatolian High School in 2001, he got admitted to the Koç University where he received his B.Sc. degree in Computer Engineering. At the undergraduate level his focus was on 'Software Verification' and 'Quantum Computation'. He enrolled the program of Master of Science in Electrical and Computer Engineering at Koç University in 2006. During his years at the graduate school, he studied language change, language evolution, computational and cognitive linguistics. In all the fields he submitted his work to various conferences.

He enrolled Phd Program in Cognitive and Linguistic Sciences at Brown University in September, 2008.