Rijks*universiteit* Groningen

Academiejaar 2005–2006

THE EMERGENCE AND MAINTENANCE OF LINGUISTIC DIVERSITY OUT OF A UNIFORM LANGUAGE POPULATION

author
Eelke van FOEKEN

internal advisor
Bart de BOER

second reader
Jennifer SPENADER

955
2006

Contents

Chapter 1

Introduction

In Genesis Chapter 11 of the Bible, the Tower of Babel was the tower built by a united humanity to reach heaven. In these days all people spoke a single language. God came down on earth and saw what the people were doing. He feared mankind, having one language, could become capable of doing anything and threaten the power of God. To prevent this from happening he confused their languages so that everybody spoke another language. The builders were then spread around the world. Below you can read the relevant passage of Genesis.

Gen 11.1	NOW the whole earth had one language and one speech.
Gen 11:2	And it came to pass, as they journeyed from the east, that they found a plain in the land of Shinar, and they dwelt there.
Gen 11:3	Then they said to one another, "Come, let us make bricks and bake them thoroughly." They had brick for stone, and they had asphalt for mortar.
Gen 11:4	And they said, "Come, let us build ourselves a city, and a tower whose top is in the heavens; let us make a name for ourselves, lest we be scattered abroad over the face of the whole earth."
Gen 11:5	But the LORD came down to see the city and the tower which the sons of men had built.
Gen 11:6	And the LORD said, "Indeed the people are one and they all have one language, and this is what they begin to do; now nothing that they propose to do will be withheld from them.
Gen 11:7	"Come, let Us go down and there confuse their language, that they may not understand one another's speech."
Gen 11:8	So the LORD scattered them abroad from there over the face of all the earth, and they ceased building the city.
Gen 11:9	Therefore its name is called Babel, because there the LORD confused the language of all the earth; and from there the LORD scattered them abroad over the face of all the earth.

If the core of this project was that God and nothing else has caused the diversification of language it would be finished now. And of course it is not. But this example nicely introduces the research project, by showing the great fascination that mankind has always had with the existence of linguistic diversity.

This research is about the essence of language diversification and preservation. How do languages diversify into a variety of dialects or languages? There are many works on the subject of language change, evolution, convergence and divergence. How did language evolve into human existence? Originated language capability of humans at a single location and diffused afterwards or did it invade at several locations before diffusion? What are the mechanisms by which languages operate over human societies? How do languages change and what factors play a part in changing? These are some of these big questions that relate to this research.

1.1 The Aims

Although much research has been done on the mechanisms of language evolution and divergence, there is no insight in the minimum requirements that are necessary to cause linguistic diversity and subsequently stabilize it. The focus in this project therefore lies in finding the minimal requirements for a society possessing one language to evolve, or emerge, into several groups with several languages and for linguistic diversity to stay stable, thus not to fall back into convergence or diversify into for example individual languages. The first and most important part of the research question is:

Research question part I What are the minimal requirements for first, causing a group of agents with one language to diverge into several reasonably large groups with more languages, and second, for preserving the achieved linguistic diversity?

In order to answer this research question a Spatial Population Model (SPM) has been developed, which simulates a space on which individuals live and transmit language to their children. The model should provide new insights in the minimal mechanisms necessary by which a single language diverges into several languages and how the linguistic diversity is maintained.

The representation and the mechanism of transmitting language in the SPM are made very general and simplistic and the representation is not based on a specific language, This is because the project is about linguistic diversity in general. Although the representation and mechanism are simplistic they are realistic. They are based on general features of language. The mechanism provides the possibility for the development of new languages. In models that are similar in its nature, the simulated languages (e.g. [Livingstone and Fyfe, 1999/ Steels and McIntyre, 1997/ Nettle, 1999]), are not well theoretically based; they are quite ad hoc. The second important aim in this thesis is therefore to develop a representation and a mechanism of transmitting language which have a theoretical foundation. This will make the results and conclusions stronger. The representation and mechanism of language transmission are based on a modern theory of language, Stochastic Optimality Theory (SOT) [Boersma, 1997], which has never been used in similar models.

Research question part II Will the application of the modern theory of language, Stochastic Optimality Theory, work properly and result into new conclusions

1.2 The Contributions

The main theoretic sources for this thesis are: *Artificial Life*, and *Stochastic Optimality Theory*. Artificial Life is an important field of research within Artificial Intelligence and is used for the modeling of biological multi-agent interacting systems. The use of multiple autonomous agents communicating in a simulated environment is part of the field of autonomous systems, a subdivision of the master Artificial Intelligence, but specifically in the subfields of language evolution and multi-agent systems. In this thesis the aim is to contribute to the field of Artificial Life by developing a new Artificial Life model that is able provide answers about the influence of mechanisms of linguistic diversity. Knowledge from anthropology, language theory (linguistics), biology, programming, multi-agent systems and psychology are combined and make it a multi-disciplinary project.

SOT is a language theory which is an extension of Optimality Theory (OT) developed by Prince and Smolensky [1993]. This thesis aims to contribute to SOT. This new representation and mechanism of transmitting language will hopefully result into new conclusions about SOT (for example that SOT can be used in a very general and abstract way to describe language).

Language is obviously an important part of this thesis, but how is language related to Artificial Intelligence? Language is such an important discipline within Artificial Intelligence because language is the feature that distinguishes us from other animals, and AI claims language is an important part of intelligence.

1.3 Background

This research project about linguistic diversity will be in the line of the research done by Bart de Boer, who is specialized in language evolution. The subject is new within AI in Groningen and this new model could be the base for more research.

In the SPM there is need for a representation of language that is biologically plausible: this language needs to be transmitted to or learned by children; this language is an abstraction of human language and encompasses the grammatical, the phonological, the lexical and the cognitive aspects of language. The background information for this thesis comes first of all from Boersma [1997], who developed the SOT, and is a language theory which is biological plausible. Understanding this language theory is important for thesis, because the language model is largely based on it. Boersma developed a representation of language and an algorithm for learning the language called the *Maximal Gradual Learning Algorithm* (MGLA). SOT is suitable for this project, because it can account for first, the transmission of language with a certain variation or the *discontinuity of language transmission*. The transmission of language to children needs to be imperfect, otherwise the languages would stay the same throughout the simulation. This imperfect transmission from adult to children stands at the basis of language change in this model. Second, the language needs to be able to innovate, so that new languages can emerge. This also can be done by using SOT. Third, this theory is never used as a representation of langauge. And last, the developers of Optimality theory aim to have a theory that can account for all aspects of language. This is needed here because the language model is an abstraction of all aspects of language.

These many aspects come forward on different scales. For example on the smallest scale, there is language variation within individuals, and on a large scale, over societies, with different languages. In between scales are social language groups, dialects, etc.. Because this model is an abstraction of multiple aspects of language the discussion in chapter 2 and 3 are about this.

In this thesis the theory of langauge is based on the conviction that language is socially and culturally transmitted and that the biological evolution is also socially and culturally driven. The reasoning behind this conviction comes from de Boer [1999]/ Elliot [1995]/ Nettle [1999]. A huge source of information on linguistic diversity has been Daniel Nettle ([Nettle, 1998] and [Nettle, 1999]), who has provided information about the definitions of linguistic diversity, striking examples of linguistic diversity, several mechanisms of language evolution and of a model he developed which is covered in this thesis. Further information about mechanisms of language evolution come from Luc Steels ([Steels and McIntyre, 1997/ Steels and Kaplan, 1998/ Steels, 1998/ Livingstone and Fyfe, 1999/ Trudgill, 2000]). The background on how the development of learning language by children proceeds is quite important in order to support the choices made in the model. Information on this has mainly come from Sethuraman [1996].

1.4 Models

In this project two models have been developed: The main model is the *Spatial Population Model* (SPM) and the other model is the *Individual Language Model* (ILM). The SPM resembles a simulation model made by Nettle in his book Language Diversity [Nettle, 1999] in that there are multiple agents on a grid which possess a certain language which are transmitted to children and that the evolution of language in a population over the generations can be tracked. But the SPM is different in the sense that a new theory of language and language transmission is used. The SPM is a complex dynamic system simulating a 2D environment with individuals that possess and transmit language to children. This 2D environment is divided into discrete sites. Multiple individuals or *agents* occupy this discrete grid. These agents possess a language. Over time the agents grow older and reproduce. Generations follow up generations, while the languages in the population change or evolve. Essentially, every instantiation of this model starts with a group of agents, distributed over a space, which start with a uniform language. This model is used to investigate the development of the language diversity over time. Because multiple parameters can be adjusted, the space of possibilities for investigation is large. Consequently, with this model it is possible to perform a thorough search in finding the minimal requirements for causing linguistic diversity and stability. The language divergence is a matter of emergence out of local interactions. Because emergence entails unexpected global events, there is a lot of tweaking with parameters and patience necessary to get preferred results.

The ILM is developed for ensuring that the mechanism of language transmission between an adult and a child works well. Furthermore, the ILM is used to test the reaction on several parameters. Also, the ILM is used to test the evolution of language over several generations of single adult and child agent language transmissions. Finally, with the tests done on this ILM, conclusions can be made about the quality of the application of SOT. In the SPM, there are multiple language agents, which have amongst themselves, multiple language interactions at every step of a simulation run. The individual language and the language interactions between adult and child are highly important, because they are the basic instruments. Consequently, it is important to start with a well tested individual language model (ILM), which is guaranteed to work well between a single child and a single or adult. Thereafter, the language and method of language transmission, tested in the ILM can safely be implemented and applied in the *SPM*. In order to achieve these goals, this separate ILM has been developed and tested. This model is without spatial dimension.

1.5 How to Read

The thesis starts with two background chapters. The first chapter contains an overview of language and its evolution. This starts with an overview of language, how it relates to this thesis and what theory of language transmission and evolution is supported. Also the definition of linguistic diversity is introduced. Then there is a complete overview of the mechanisms of language evolution, of which some are tested in the experiments. This is followed by a description of a model of linguistic diversity developed by Nettle [1999], which has close relations with the model in this thesis. The last section is about a model of cultural diversity developed by Axelrod [1997], which has important similarities with linguistic diversity. In the second chapter an overview of the SOT is given, which starts by introducing OT, following an elaborate description of the representation and learning mechanism (MGLA) of language in SOT.

After the background chapters comes the description of the behavior of the two models, the Individual Language Model and the Spatial Population Model. Then there is a chapter about the measures that have been used during the experiments. Especially the importance of the

Hierarchical Social Entropy measure in this thesis has earned itself an elaborate description. Then there is a chapter on the several visualizations and graphs of the model's simulation.

The results of the experiments with the *Individual Language Model* are treated in chapter 6. Here the influence of several parameters on the language acquisition of children are quantitatively and qualitatively determined. Also, the development of language over a single generation line of agents is tested by using the *Iterated Learning Model* ([Kirby, 2001]). Thereafter, several basic parameter settings for the SPM are determined based on the precluded results. In chapter 7 the results of the experiments with the SPM are shown and discussed. First there is a general description of the development of the model over the generations to give an intuitive feeling. The main part of the chapter consists of the presentation of the results of, first, the tentative exploration of the influences of the several involved parameters and second the sensitive study which provides an aimed research in finding the minimum requirements necessary. All results are presented in graphs with different measures, displaying the 1st 2nd (median) and 3rd quartiles over time. Finally there is a discussion of the results. The last chapter of this thesis is of course the conclusion.

Chapter 2

An Overview of Language and its Evolution

This research is, speaking in very general terms, about linguistic diversity in a population of individuals and the mechanisms of language change. The important concepts are *language*, *linguistic diversity*, and *mechanisms of language change*. This chapter is an overview of these concepts. First the concept of *language* is introduced, and the relationships between this concept and the language model used in this project. Second, a hypothesis is made about how language is represented in humans and how language evolves. This is explained in section 2.2. Third, it is important to have a definition of linguistic diversity, which is given in section 2.3. Fourth, the origin of language can teach us about how language spread and diverged into the world and may support an assumption made in this research so this is discussed in section 2.4. Fifth, because the main aim of this research is to search for the minimal requirements to cause linguistic diversity, it is important to have an overview of the mechanisms of language evolution.

2.1 Language

The model used in this project uses several aspects of language. It is therefore essential to have a notion of what language is and to know which aspects of language are represented in this project.

Saussure [1987] gives a complete view of language and claims language consists of three distinct, but inter-related mechanisms:

1. *langage*, the physical, cognitive and cultural bases for spoken language
2. *langue*, the lexical, phonological and grammatical structures of a particular language
3. *parole*, the actual speech produced by a particular individual.

How do these three mechanisms relate to the model of language used in this project? The first mechanism, *langage* relates to this project, first, because language is transmitted *culturally* and *socially* to children, and second, because the physical and cognitive bases is reflected as a cause of the *discontinuity of language transmission*, see section 2.6.1 for a more detailed discussion. The second mechanism, *langue*, relates to this project because the representation of language in single agents is an abstraction of the lexical, phonological and grammatical structures of language. The third mechanism, *parole*, relates because the language is transmitted through the production of simplified speech by adults to children. The word *particular* is important because it signifies the imperfect language that individual speakers actually produce, with speech errors, reductions and interruptions.

The *langue* of a language can be used to compare different languages and consequently to determine the degree of similarity or difference between them. In other words, languages can differ on three levels: the *vocabulary* or words, *pronunciation* or sounds and the *grammar*. *Accents* of a single language only differ in pronunciation, where a *dialect* differs on all three aspects. In this project it is important to have an idea of how to compare languages, because it is necessary to divide the language population into several levels of similarity, and eventually label these levels into, for example dialects or accents.

One of the main problems for this project, is to determine when two vernaculars are considered two different languages. Surprisingly, a main observation in many societies is that political and cultural factors are more important than the, more intuitive defining characteristic, intelligibility, [Trudgill, 2000]. The difference between languages and dialects is not assured. Different national languages may possess a high degree of mutual intelligibility, while so-called 'dialects' of one language may be relatively unintelligible [Livingstone and Fyfe, 1999]. There is often a *geographical* or *social dialect continuum*. This means that there is not a clear break between dialects and it means that the linguistic characteristics of these dialects change gradually. These continua make the division between dialects and languages even harder, because there exist areas where the neighboring areas have mutual intelligible languages, but dialects at distant points are unintelligible [Livingstone and Fyfe, 1999]. An example is visible around the border between Germany and the Netherlands. The Dutch population near the German border can understand and talk with German inhabitants. The languages are quite similar, while Dutch and German are officially considered two different languages. In this project two vernaculars are considered two different languages when they are mutual unintelligible.

Trudgill [2000] observes there can be several languages within a country, but there can also be one language spread over several countries . Trudgill calls the last a *superposed variety* of language. An example is Standard English of which the grammar is the same in America and Australia. Standard English is a language mostly spoken within the higher educated groups. Standard English does not have a standard pronunciation or accent. Trudgill [2000] gives RP as an example of a high-class language where the pronunciation and the accent are the same, no matter where. These two examples show that languages may or may not differ in place, in grammar, accent or pronunciation.

An important aspect of language is the way in which it is transmitted from a competent adult speaker to a learner, mostly a child. Is this done genetically or culturally and socially?

2.2 Innateness of Language versus Cultural Evolution of Language

It is important to have a concept about how the transmission of language to a child works and on a more global level how the evolution of language works, because the model needs to be based on this conviction. Essentially there are three issues which arise when studying language and brain development, which are *innateness, modularity* and *domain specificity*. *Innateness* refers to the idea that language is acquired via a set of principles at birth. *Modularity* means that language is a separate cognitive ability, which is dependent of the development of other brain structures. Lastly, *domain specificity* means that there are specific parts of the brain which are concerned with language learning and processing.

Generally, there are two extreme views on language and brain development. *Functionalists* claim language to be non-innate, non-modular and not domain-specific. *Nativists* claims just the opposite and think positive on innateness, modularity and domain-specificity. It seems likely that some middle ground between there views is most plausible. The second view is supported by Chomsky [1995] and is explained well by Pinker [1994]. Pinker accepts Chomsky's account of language, where the concept of *universal grammar* is central. The essential feature of the universal grammar's present form is termed by Chomsky the Principles and Parameters approach. This approach entails that the underlying structures of language, the grammar, are innate and the same for all humans and different languages are the result of ascribing binary values to a small set of parameters, [Chomsky, 1995]. Pinker supports the possible evolution of universal grammar.

However, there are several problems with innateness and the evolution of an universal grammar, which are mentioned in de Boer [1999]. First, there is no general accordance about what the universal grammar would look like. Second, the growth of the brain is too unspecific to create something as specific as a universal grammar. Third, there is evidence that children are highly flexible in recovering from brain damage. Children are capable of learning language flawlessly after areas in the brain, associated with language in humans, (e.g. Broca's area and Wernicke's area) are severed. This signifies other regions in the brain can adopt the language functions. It is not logical that a very specific universal grammar suddenly is able to appear in another area of the brain. Fourth, there are problems with the evolution of an universal grammar. The evolution of an universal grammar should entail that some group of individuals needs to have a clear survival benefit from a variation that has mutated in the universal grammar. This clearly seems unconvincing. It has to be a group of individuals , because language is a social phenomenon. And to cause a whole group to have a slight variation of language, every individual in the group needs to have nearly simultaneously evolved the same variation, which seems highly unlikely. The final and perhaps most important error of Pinker according to Elliot [1995] is the mistaken view of natural selection that a complex physical structure serving a specific function is constrained to change only gradually. To quote from Elliot [1995]:

> Natural selection also operates through serendipitous transfer of complexity developed for one function to a new function, typically the move from swim-bladder to lung, from webbed foot to wing, from gill to structures of the ear and so on.

Moreover the spreading and emergence of linguistic innovations is much faster [Nettle, 1999] than would be when biological evolution is the only mechanism of language change. The success of a linguistic innovation depends on its adoption by the whole group. It does not depend on the single mutated individual, because this doesn't make it a better individual. Theorists arguing against the innateness of language do not need to explain why there are universal similarities in the languages of the world. de Boer [1999] shows universal features of languages can be explained without appealing to innate structures . Self-organizing interactions in a population are sufficient to explain the emergence of structure.

In line with the hypothesis of Bart de Boer in his thesis [de Boer, 1999] and supported by Trudgill [2000], Nettle [1999], Livingstone and Fyfe [1999] and Steels and McIntyre [1997], and for the above mentioned reasons a more functionalist view is supported here and means that language is not innate, but that language is culturally and socially transmitted to children and that the biological evolution of language is culturally and socially driven. The last part of the hypothesis means that biological adaptation to language only became advantageous, after language became more complicated through cultural mechanisms and other mechanisms of language change. It has to be noted that the view does not necessarily denies domain specificity.

The primarily social character of language is also advocated by Nettle [1999] and Trudgill [2000]. Nettle thinks the social selection is a very important influence on language transmission. Social selection means that a learner prefers a certain language group over another because of social reasons. Sociolinguistics is concerned with the social and cultural aspects of language. Trudgill

[2000] says social attitudes can partly explain dialect changes. There exist so called *sociolects* which reflect the social class in which it is spoken.

Some parts in this section already touched on the concept of linguistic diversity. In the next section this notion will be introduced in more detail.

2.3 Introducing Linguistic Diversity

What is linguistic diversity? A citation from Nettle [1999] gives a sufficient summary of the range of linguistic diversity.

> Humankind today speaks about 6,500 different, mutually unintelligible languages. These languages belong to at least 250 identifiable large families, though there are various proposals to group these into still larger units. Within these families, there are languages that use a dozen contrastive sounds, and languages that use 100. There are languages that place the subject of the sentence before the verb and languages that put the verb first. There are a few that place the object before either. Some languages mark the relationships between the constituents of the sentence, or between the sentence and the world, by extensive inflection, whilst others use almost none, and rely on independent particles and the order of words.

Nettle [1998] classifies three types of linguistic diversity:

Language Diversity Regions can differ in the amount of different languages these hold. For example, Papua New Guinea has 862 languages listed in the Ethnologue [Gordon, 2005], which makes it linguistically the most diverse place. This is a simple measure counting the number of languages.

Phylogenetic diversity of languages Nettle claims this measure of diversity needs to be distinguished from language diversity. It is an indication of how many families or branches of families are present. He mentions the Papua New Guinea example where the language diversity but also the phylogenetic diversity of languages is high. But the two types do not necessarily go hand in hand. In Central Africa there are many different languages, making the language diversity high, but they are all closely related and all belong to the Bantu family, making the phylogenetic diversity low. An opposite example are the languages in Latin America. In this continent there are little different languages, making the language diversity low. But these languages belong to dozens of families, making the phylogenetic diversity high

Structural diversity on some linguistic parameter Languages can differ on several linguistic parameters. For example, languages can differ in their basic constituents Subject, Object and Verb. There exist SOV, SVO and VSO languages. A set of languages has high structural diversity in word order, when there are many different word orders in the different languages. Structural and phylogenetic diversity will often correlate because many different families usually entails many different structural types of language.

These three measures of linguistic diversity are used in some way in this project to measure the linguistic diversity in the simulations. See section 5 for how these measures are incorporated in the model. The model in this project has a measure for language diversity by a simple count of the number of languages in the whole population. Another measure called *entropy*, combines the language diversity with the proportion of all languages. The *Hierarchic Social Entropy* (HSO) measure is related to the phylogenetic diversity and to structural diversity, because the degree of difference between languages is incorporated, indicating that a population with very different languages (more families) has higher HSO than a population with the same amount of but more related languages. The difference between languages is measured by comparing the order of the elements between the simplified languages. The theory in this project is that the order of the elements in the language determine the output of the individual, for example the subject-verb-object order. As one can see this relates to structural diversity. It is important to note that the languages in this project are very simplistic, but it is delightful to notice that the several types of linguistic diversity applied on real languages are also applicable on the languages in this model. In this project it is important to note the scales at which the languages can differ, to later describe the languages of the population in the model in terms of, for example, families or dialects.

There are some striking examples which illustrate the complexity and fascination of linguistic diversity. Papua New Guinea is one of them, where there are so many languages on an island with only 5.6 million inhabitants. A simple calculation results in that on average there are 6500 native speakers for every language. Another example is in Philadelphia where Black English dialects are diverging from standard US English, despite the day to day interaction with normal English talking people and television (Labov, cited in Chambers [1983]). What are the mechanisms by which these languages diversify?

Modern language most have had a starting point. Maybe, here is something to be learned from. In order to know more about how language diverges it is interesting to investigate the origin of language and how language diverged then.

2.4 The Origin of Language

With archeology, biology and history of humankind it is possible to track the development of languages over the centuries. Specifically, maybe something can be learned about how language originated and how language diversified or converged. An important question in the evolution of modern language is whether language capability arose at a single location and diffused afterwards (monogenesis) or invaded at several locations before diffusing (polygenesis) [Wang and Minett, 2005]. Wang and Minett [2005] say that it is likely that modern language first occurred in human behavior between 160000 and 50000 years ago. The first fossilized remains of modern humans that are found date 160000 years back and it is widely believed that modern language had already emerged at the time of the *cultural explosion*, which is around 50000 years ago. So lets assume modern language arose 120000 years ago. Assuming monogenesis, which is widely believed, it started with a single language and must eventually have diverged into multiple languages. Somehow a uniform language population diverged into several language populations. The monogenesis hypothesis supports the possibility of divergence of languages out of a single language, which is an assumption of this project.

2.5 Language as an Open, Complex Dynamic System

Before the mechanisms of language change are discussed it is important to know that language is a complex dynamic system and an open system. This information is retrieved from the background chapter in de Boer [1999] and also largely follows the line in this part. Language is an open system, because of two things. First it is an open system, because individual speakers can enter and leave a language community without changing the language. Second, it is an open system because innovations of linguistic structures can leave (disappear) and enter (innovate) the language. Words or grammatical structures can arise or disappear. As de Boer indicates, it is necessary to have more insight in these phenomena like language change and linguistic diversity, by looking at language as an adaptive and complex dynamic system.

A *complex dynamic system* consists of a large number of only locally and interacting elements. There is not any global interaction. The local interactions are non-linear and non-hierarchical, which means that the global behavior or behavior of the whole system is nonpredictable. The fascinating thing about these systems is that there can occur or emerge global organization without global interactions. An example of a complex dynamic system is the flocking behavior of birds or fish. There is a globally organized group of these creatures without any creature with higher cognition which is able to organize this grouping. The flock has a certain direction and a certain distance between individuals. There is no central authority. The only three rules that are necessary for every single creature to adhere to in order to create global flocking behavior is first, aligning with neighboring creatures, second, separate from crowding creatures and last, move to the average position of local flockmates. Language is also such a complex dynamic system according to de Boer [1999]. de Boer believes the global organization of language expresses itself in the remaining coherence of language. This global organization is emergent because there are only locally interacting individual language users (they talk to each other and learn the language from each other) and there is no central authority controlling the language.

In this project this view of language being a complex dynamic system is supported. Therefore, the model consists of a set of locally interacting agents whereout globally organized language behavior may be observed. There is no central authority. Secondary, the global spatial organization of the agents can also be seen as emergent, because the agents only individually move with a small probability to another place. Although there are no local interactions between agents, the movement of a single agent can be seen as local. The global behavior is that the agents are reasonably equally spread over the spatial grid.

2.6 An Overview of the Mechanisms of Language Evolution

It is important to explore the theory of language evolution and specifically the basic mechanisms which underlie language evolution, because these mechanisms can be applied in the model of this project. The goal is to find minimal set of mechanisms necessary to cause linguistic diversity and maintenance of the achieved diversity. Several authors have made up categories in which to divide the mechanisms of language change.

Nettle [1999] observes that language evolution and divergence occurs because of two types of mechanisms: *sources of variation* which cause a change in language, with discontinuity of language transmission as the main one and only one mentioned in this overview, and *amplifiers of variation*, which are mechanisms for increasing the small differences caused by sources of variation. Another categorization is made by Steels and McIntyre [1997] who claim that languages can change either *internally* or *externally*. An internal cause may be the propagation of a new sound causing a chain shift of changes in existing sounds. External causes are geographical isolation or social barriers. In this project the categories sources and amplifiers are used.

In this project three mechanisms of language evolution have been tested: Discontinuity of Language Transmission (DLT), spatial organization and mobility. These concepts are most important and are covered first. Thereafter all the other possible mechanisms are briefly noted, but because they are not used in this project they are not explained extensively.

2.6.1 Discontinuity of Language Transmission, a Source of Variation

Nettle mentions Meillet [1926], who claims that linguistic variation arises because of the *Discontinuity of Language Transmission* (DLT). This mechanism of language change is considered a source of variation, because it can cause a change in the language. There are two ways how DLT can occur, which both can be seen as internal causes. First, there is the way from adult-speakers mapping their *linguistic competence*, which means how the language is represented in our head, into *linguistic performance*, the actual finite set of utterances produced in real-time communicative situations. The *linguistic performance* is influenced by cognitive and physical systems, and will have discrepancies with the *linguistic competence*. Speakers have their own idiolect, or personal and incomplete knowledge of the language. Throughout the thesis this first stage of DLT is called *imperfect performance*. Second, a child-learner will be exposed to a finite sample of linguistic performance, which comes from different speakers with different linguistic performances and the child must interpret this data into the linguistic system. This reverse mapping does not proceed without mistakes and results in *imperfect learning*. Throughout the thesis this stage of DLT is called *imperfect learning* These two stages logically result in variation. Steels and McIntyre [1997] gives a nice example of discontinuity of language transmission. Language learners may interpret some linguistic behavior as rule-governed, while it is not. When learners use this in the next generation as a rule this is a linguistic innovation.

In this project there is need for discontinuity of language transmission in children. This is achieved by using both ways of DLT: imperfect performance and imperfect learning.

2.6.2 Spatial Organization

A intuitively recognized cause of language change is the spatial organization of language. This means that the distances over which agents can communicate are limited. One can imagine that individuals, and then especially child-learners, only have a certain limited amount of people they communicate with or learn from. These people are mostly limited by distance. Two totally separate spatial locations that do not have direct interaction, can be imagined to have different language drifts over time. Eventually these two languages will be different. In the model in this thesis, spatial organization is incorporated, because the agents are placed on a 2D grid. However, the degree of spatial organization can be limited by increasing the distance over which children can learn from adults.

Models on Spatial Organization

Several models have been developed, where the influence of spatial organization on language has been investigated. In [Livingstone and Fyfe, 1999] a model is made where the results show that spatial organization has significant effect on linguistic diversity. Two versions of a simulation model are made. The second implementation incorporated spatial organization, limiting the distances over which agents could communicate. The general observation made here is that global language diversity was high but local language diversity stayed low, due to this spatial organization. This maintenance of low local but high global language diversity is also one of the goals in the model, coping with the second part of the research question.

Steels and McIntyre [1997] focusses also on language change in the communication structure between linguistic communities determined by spatial location. Here the naming of objects with the aim of developing a common lexicon is investigated. The agents are spatially distributed to observe the changes in the community's structure on language. Naming games are especially useful in the creation, transmission and evolution of linguistic conventions. The naming game is a game where a single speaking agent identifies an object by using a name. The game succeeds if the hearing agent agrees on the name for the object. The naming game is adaptive when both agents are able to change their rules. The naming game is very simplistic and many issues, like the emergence of ambiguity, multiple interaction levels, the grounding in the real-world sensing, or actuation, cannot be studied. The hearer adopts the name of a object to the speaker's name of the object if the naming game failed. This paper is interesting because it also uses spatial location and because a very simple language structure is used.

In the paper of Steels and McIntyre [1997] the communication process is reliable and unambiguous, but another study by Steels and Kaplan [1998] shows that conventions may both emerge and continue to exist under conditions of imperfect communication. This latter study relates more to this study because imperfect communication is also used. A main observation is the sigmoid shape that can be seen in the learning curve of communitive success, with and without spatial organization. In the spatial case a stable language develops within clusters. But a second language, *interlingua*, which is weaker will be shared among different clusters. The stronger the interaction between clusters the stronger the interlingua. This spatial organization therefore causes language diversity. The divergence or coherence can be tuned by changing the probability of interaction between communities (clusters). If interaction increases the communication between clusters decreases, consequently increasing the differentiation. When communication between clusters increases, coherence increases.

From these examples it generally can be said that increasing distance and decreasing the amount of communication between groups of individuals on different spatial locations increase the amount of linguistic diversity.

2.6.3 Mobility and Geographical Isolation

The degree of interaction between groups of individuals can also be influenced by the amount of *mobility* between these groups. Higher mobility leads to more interactions, and probably leads to less linguistic diversity. An important question relating with this project is:

> Why do some linguistic innovations spread faster than others?

Factors which play a role are if the innovations are spread from person to person, group to group or by traveling persons. It depends on the mobility and mingling of people in a society. Differentiation of dialects will emerge faster with less geographical mobility. More geographical mobility will lead to dialect leveling. When there is no interaction between groups this can be seen as complete isolation. This can be caused by *geographical isolation* [Nettle, 1999], for example rivers and mountains etc., which causes populations to be isolated from each other, having no interaction and therefore causing drifts of random variations in language in different directions. Natural barriers, which cause geographical isolation, are for example rivers or mountains and geographical distance. In a situation where two groups are *geographically* separated, and because the variation for every new generation is random, the languages in the two groups will have a large chance of drifting apart as time progresses. Eventually these languages can even be completely different, or at least mutually unintelligible.

But even in situations where there is mobility, with interactions with other societies, linguistic diversification occurs.

2.6.4 Other Mechanisms of Language Evolution

DLT, spatial organization and mobility are the mechanisms which are used and tested in this project. There are several other mechanisms of language evolution which are important to know about in order to have a more complete overview.

First there are social influences on language. *Social selection* is one of the mentioned biases or amplifiers for language diversification in Nettle [1999]. Moreover, in this thesis language is considered to be socially and culturally transmitted. Language has a primary social character.

> **Social Selection:** A learner prefers a certain language group over another because of social reasons.

This means the learner is selectively biased toward a certain variant of language. Social reasons can be class, sex, age or religion. Social attitudes towards languages can partly explain dialect changes [Trudgill, 2000]. With the concept *Social selection* the realm of *sociolinguistics* is entered, which is discussed by Trudgill [2000]. In citetLa63 two examples are mentioned which show that linguistic diversification can result without geographical isolation.

Second, there is *functional selection* which is an amplifier of variation and an internal mechanism of language change Nettle [1999]. Several mechanisms of language change which relate to functional selection are mentioned by de Boer [1999] and Steels and Kaplan [1998]. Children perform so *functional selection* [Nettle, 1999] on language, which means that they unconsciously learn another linguistic structure above another because of a lower processing load. It means that a phonological form that is less easily picked up or heard will be harder to process than another; the last will be learned. An easy to parse or remember grammatical form will be more likely learned, in comparison with a very complex grammatical form expressing the same meaning. When a sound changes in a language, this may have functional consequences, where other successive sounds are more difficult in combination with the new sound. The linguistic process of functional selection is analogue to natural selection. Language evolution tends to creating languages that are more efficiently learned and spoken.

Third there is *cultural evolution* de Boer [1999]. Cultural evolution means that not the individuals but the ideas or knowledge of a population undergo evolution. In case of language this is the knowledge of the language. The knowledge of language is transmitted through learning by new speakers from old speakers. Selection occurs through the criteria of the before mentioned functional constraints. And variation or language change occurs through the DLT and by conscious innovation, for example the invention of new words because of new cultural items.

2.6.5 Summary of the Mechanisms of Language Evolution

In this section several mechanisms of language evolution are treated. The basic mechanism, discontinuity of language transmission, is the basic mechanism used in this project. Another mechanism that is incorporated initially is the spatial organization. The model is a two-dimensional grid of several locations. Consequently, there exist distances between agents on different sites, which limit the possibility of communication. Geographical isolation and mobility are also two possible mechanisms in the model. Global organization occurs automatically in this model. Agents locally communicate, from which a global pattern of linguistic diversity occurs. The movement of the agents occurs locally, resulting in a global pattern of equal spread of the agents. One has to note that the self-organization of language does not occur under functional constraints in this model. Functional, social selection and cultural evolution are both mechanisms that are not incorporated in the model initially but are optional for further research.

2.7 The Model of Language Diversity developed by Daniel Nettle

This section treats a language diversity model developed by Nettle [1999]. This discussion of this model touches several important problems which are discussed next. Also simulations are tested on several mechanisms of language evolution, and their influences.

2.7.1 The Neutral Model

As a starting point Nettle [1999] discusses the *neutral model*. This neutral model, as will be shown later, is quite similar to the initial instantiation of the model in this project. This model is partially based on the neutral-mutation model of biological evolution [Kimura, 1983]. Here biological diversification is caused by the small rate of random genetic mutation and geographical isolation.

The neutral model has a different method in the transmission of genes and linguistic features. In the neutral model, language is learned by children who, when they become adult, teach their language to the next generation. The language transmission is not perfect. Every new generation of speakers learns a different minor random variant of a language, so there is a Discontinuity of Language Transmission.

The conclusion is that the path of linguistic diversification is a random one. Generation upon generation, a random minor variant of the language is learned, making it a random path. In our model the children also learn a small random variant of the language. The variation resulting from production and acquisition is concluded to be sufficiently unpredictable for the neutral model to hold by Nettle [1999].

2.7.2 Key Problems for the Neutral Model

The neutral model has several serious problems and Nettle questions if it is an adequate account of language evolution. Nettle mentions four key problems. These problems are important, because these provide insight in the nature of language and language learning.

The Averaging Problem

The *averaging problem* is: children learn language from the people, social group, around them (see section 2.6.2). One hypothesis is that the child learns the average of the language around him/her, using some kind of error-minimizing strategy. Consequently, eventually the random changes (variation) that appear in the production and acquisition will even out, leading to no change. In biological evolution, and therefore also in the neutral model, random mutations are inherited by the children. The hypothetical averaging feature of language makes it largely impossible for a single mutation to remain in a population.

The Threshold Problem

According to Nettle it would be necessary for the neutral model of language to be continuous. An example of a *continuous* linguistic variable is the vowel gradation, which is used in the computer simulation, discussed in 2.7.3. But also according to Nettle, most linguistic variables are not *continuous* but *discrete*. *Discrete* variables consist of the choice between definite rules, either this or that and nothing in between. For example; languages differ in the order verb-subject or subject-verb. When verb-subject order is prominent in a language but sometimes, due to some source of variation the subject-verb order arises the learner of a language will most of the times adopt the majority form verb-subject. This is called the *threshold problem* and is the second problem. So when a variation on some *discrete* rule arises, because of the rarity of occurrence, this variation will rarely be passed to new learners.

Solution for the Averaging and Threshold Problems

The *averaging* and *threshold* problems can be overcome when the learner is selectively biased towards certain kinds of variants. This can be done by incorporating the biases *social* and *functional* selection.

Two Other Problems

The third problem for the neutral model is that diversification sometimes also occurs in absence of geographical isolation (section 2.6.3). With the neutral model this is difficult to account for. This problem is directly related to the research question about how linguistic diversity emerges out of a single language group and how the achieved language diversity is preserved. In this project an attempt will be made to establish linguistic diversity without geographical isolation. The fourth problem is that the path of linguistic diversification is not a random one, as the neutral model predicts. As Nettle points out there are patterns of structural correlation in the world's languages that represent parallel evolution.

Nettle thinks that the first three problems can be overcome within the neutral model, but that the fourth requires an extension or change. The key problems are summed up here, and it can be concluded that the neutral model is probably an uncomplete theory of language diversification. It is probably true that, as the neutral model assumes, language is imperfectly transmitted from adults to child. But there are additional influences on the transmission of languages over generations, that need to be considered to account for the problems mentioned here. This is in contrast with the neutral model where the learner has not any biases, but simply averages over everyone.

2.7.3 Computer Simulation: chapter 3 from Nettle's Linguistic Diversity

Nettle [1999] supported and tested his ideas by doing computer simulations. These are important to describe, because the questions he tries to answer by running the simulations are very similar to the ones this project, and because the model in this project will be starting with similar initial conditions.

This model Nettle is trying to investigate solutions for the four key problems mentioned before. The earlier mentioned, 2.7.2, *averaging* and *threshold* problems pose problems for groups of a realistic size. The third problem Nettle points out is that the neutral model is unable to account for the finding that linguistic boundaries can develop and persist in absence of geographical isolation. The fourth problem is that the path of linguistic diversification is not a random one, as the neutral model predicts. There are patterns of structural correlation in the world's languages that represent parallel evolution.

The model of Nettle starts with a simulation where initially the members of the population share the same language. Next, several mechanisms are applied to study the effect on language in time. The population lives on a 7X7 grid. Every location is occupied by a group of 20 individuals. The grid is displayed in figure 2.1. This number is chosen because 20 is recognized as the average number of a hunter-gatherer band by several anthropologists. The individuals are between the age of 1 and 5 and at age 5 they die and are replaced by a new individual of age 1. Consequently the number of individuals stays constant. The newborn individuals learn the language from the individuals around them, which suggests a *critical period*. The distribution of age is equal, so there are 4 individuals of each age.

In section 2.6.1 the two ways, *imperfect performance* and *imperfect learning*, of discontinuity of language diversity were discussed. From these two Nettle only uses imperfect learning of the

Figure 2.1: The 7x7 grid of the simulation of Nettle

	1	2	3	4	5	6	7
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0

language learner as a source of variation and does not use variation in performance of the teacher. He claims imperfect learning is a more powerful source of variation because it reflects more the characteristics of the individuals, but in this project both are simultaneously used.

As discussed earlier in 2.7.2, in languages there is a difference between discrete and continuous linguistic items. Nettle focuses in his simulation on continuously variable items. As a set of linguistic items, Nettle has chosen the vowel system.

The child learns the language, by taking the average of all the sixteen adults around it. A noise parameter (NOISERATE) takes care of the imperfect learning by altering the final value by a random number between 0 and NOISERATE.

The initial structure consists of no communication between groups and all individuals start with similar linguistic values. In other words: the groups become completely geographically isolated. After testing in this setting Nettle adds successively inter-group contact, social selection and functional selection.

Results with Continuous Linguistic Items

When the simulation is run in the initial structure and with a non-zero NOISERATE the values of the phonemes perform a random walk through phonetic space in time. This means that each value of the phoneme, associated with a site in the grid, changes randomly over time. Every group performs such a random walk, so immediately it can be predicted that, when geographically isolated, languages have a large chance of drifting apart (language diversity). This prediction is reflected in the results.

As said earlier in section 2.7.2 and illustrated by examples mentioned in the introduction of this chapter 2.3, language diversity is capable of arising without geographical isolation, therefore with inter-group diffusion or mobility (section 2.6.3). Therefore it is necessary to account for inter-group diffusion in Nettle's model. This is incorporated by giving an individual in the first life stage a probability of value MIGRATE to swap permanently with another individual in another group chosen randomly on the grid. The individual takes the learned language with it and contributes this to the other group. Migration reduces the inter-group differences significantly and even low migration rates prevent local diversity from happening. So it seems that linguistic diversity cannot originate solely out of imperfect learning. But as will be shown in the results of this thesis this can be done with the model made in this project.

The next stage is when Nettle adds social selection. A small number of individuals within each group are randomly assigned higher status. Nettle keeps himself to the social model of LaPage [LePage, 1963], where individuals choose the language they feel identified with, consequently do not use some of the other linguistic models. In this model the individual chooses only the languages of the individuals with a higher status. When incorporating social selection, with a MIGRATE of

1 and 10 per cent, which entails 1 or 10 percent of the individuals migrate, linguistic diversity occurs. Nettle also creates a situation where in-migrants to a group cannot possess high status. This should strengthen the robustness of local norms, but this is not visible in the results.

Finally Nettle adds functional selection by using the fact that vowels tend to be evenly distributed over phonetic space. A striking example is that most languages consisting of 3 vowels have /i/, /a/ and /u/, and most languages possessing 5 vowels have /i/, /e/, /a/, /o/ and /u/. The vowels tend to be evenly distributed over the phonetic space. The results show that when there is no migration, linguistic diversity greatly increases, however with migration functional selection has no effect.

When social and functional selection are combined this produces a greater linguistic diversity than with functional selection alone. However this combination doesn't produce significantly greater linguistic diversity as with social selection alone. Nettle explains this by saying that functional selection not only increases the inter-group but also the within-group differences.

2.8 Local Convergence and Global Polarization

In Axelrod [1997], Robert Axelrod deals with a problem, which is similar to the problem in this project. The general question which he tries to answer is on page 203:

> If people tend to become more alike in their beliefs, attitudes, and behavior when they interact, why do not all such differences eventually disappear?

Axelrod proposes a mechanism that deals with how people become more similar as they interact, but also explains why convergence stops. In this project the intention is to find how a uniform language population diverges into a number of stable language regions. The difference is therefore that Axelrod starts with different groups, where this model start with a uniform group. But, the essential part which is similar is the stabilization of the number of cultural or language groups in the population.

Culture Instead of Language

Where the focus is on language in this project, Axelrod focuses on the broad concept of *culture*, which according to him is the most generic term for the things over which people influence each other. Axelrod uses a very simple notion of culture; a set of features, which each can have a number of traits.

But here the focus is on understanding the processes of culture spread, establishment or disappearance. The model of social influence proposes that human communication between similar individuals, in culture (beliefs, education, social status, etc.), is more effective than between dissimilar individuals. In turn this interaction leads to more similarity between individuals.

- The degree of cultural similarity determines the degree of interaction
- The degree of interaction determines the degree of similarity

This applies also to language, where an individual is more likely to talk to someone with similar languages then with dissimilar languages. This interaction makes the speech patterns eventually even more similar. But it is important to notice that social influence not only applies to language but also to beliefs, attitudes and behaviors.

The abstract notion of culture that is used in the model in Axelrod [1997] consists of features, such as language, religion, clothing, etc. Each feature consists of a set of values it can have, e.g. language has values German, Chinese enc. How does culture change? Are culture features interconnected in a network of traits, or are they independent? It seems it is a combination of both. The hypothesis is that some cultural features can be independently passed on to other cultures but many are embedded into larger structures of culture. Also does Axelrod, who discusses several models which treat each feature of a culture as independent.

The model Axelrod [1997] uses is new in two features. First it takes into account the effect of a single cultural feature depends on the presence or absence of other cultural features and second, it takes into account that similar individuals are more likely to influence each other than dissimilar individuals. The first feature seems unclear, because it is difficult to see how Axelrod makes the features dependent.

A main feature of the model is that there is no *central authority*. So the model is aimed at explaining cultural emergence and stability without this central authority which could have a large social influence. It only uses the convergence of culture with neighbors as a mechanism. So it does not incorporate geographical isolation, functional selection or social selection. Because culture is learned there could be a *discontinuity of culture transmission*, but there is no transmission. The model suggests a very simple mechanism and seeks to find a minimal influence which causes a stop in the convergence of culture. The mechanism is suggested to be complementary with the before mentioned mechanisms.

The model of Axelrod [1997] has a similarity with the model in this project. The similarity is that there is a geographic distribution of cultural sites. Every site is described by a variable set of cultural features. Every cultural feature can have a variable set of traits. For example: a site is described by 5 features and every feature can have 10 different traits, for example: 82330 and 67730. The basic idea is that based on the degree of similarity between two sites, the probability of interaction is determined. Therefore in the just mentioned example 40% of the features are in common, providing a chance of 40% on interaction. The interaction consists of the selection of a feature on which the active site and the neighbor differ, next equalling the active site's trait on this feature to the neighbor's trait on this feature. Interaction can only take place between immediate neighbors.

As said earlier, Axelrod claims that cultural features are embedded into larger structures; they are dependent. But the features in the model of Axelrod , in an interaction, are individually and independently transmitted.

Results

The model starts with a random distribution of cultures and proceeds by the described recipe. After initialization most sites have very different cultures, but some have a little similarity. Over time, specific cultural features tend to be shared over larger areas. The change stops when every pair of neighboring sites have identical or completely different cultures. A measure for the resistance of the social influence against complete convergence is the number of *stable cultural regions* that has developed at the end state. The number of *stable cultural regions* depend on the *cultural complexity, the range of interactions* and the *size of the geographic territory.*

Cultural complexity increases with more features and/or with more traits per feature. There where basically four results, two of which are intuitive and two are not. Important to notice here is that a model with simple rules can result in unexpected outcomes. The two intuitive results are: The number of stable regions

1. increases with the number of possible traits that each feature could take.
2. decreases with the range of interaction

The two counterintuitive results are: The number of stable regions

1. decreases with more cultural features.

2. decreases with larger territories.

Axelrod [1997] explains the first counterintuitive result by saying that the ability to interact increases with more cultural features, because there is a greater chance two sites will have some trait in common.

Other observations are that majority cultures tend to eat minority cultures even though there is no advantage in choosing the majority culture. Also, polarization occurs, even though only a mechanism of convergence for change is used. One can associate this with the majority languages swallowing minority languages

Axelrod [1997] proposes that extensions to the model can easily be made. The most interesting change for me would be to analyze *cultural drift*. Cultural drift is the spontaneous change in a trait. In this project there will be discontinuity of language transmission which is similar. Linguistic shifts are also considered as part of cultural drift. Cultural drift change continues indefinitely, therefore no region will be stable. Axelrod raises two questions which are also important for this project:

1. What is the best measure of the heterogeneity of the population when a simple count of stable regions cannot be used?

2. How long should a given run be allowed to go before the measurements are taken?

In this model it is plausible that, because of the language drift, no simple count of stable language regions can be used. In this project there is need for a good measure of heterogeneity and knowledge when the system relatively stable. Axelrod proposes two measures of heterogeneity:

1. Locally, considering the differences among all pairs of neighboring sites

2. Count the number of cultural zones or regions even though they are not stable

Axelrod also mentions two ways in which can be determined when measurements should be taken:

1. Measure after a fixed amount of time.

2. Measure after a selected measure of heterogeneity remains in a certain stable region of values.

The last way is the best for this project, because it is interesting to know how long it takes for the system to stabilize and then measure.

As Axelrod says people will likely tend to interact with others who are already quite similar to them on *relevant dimensions*. What Axelrod doesn't incorporates in his model are these relevant dimensions. He describes a culture as a string of numbers. Every feature has the same relevance. This is of course not true. Some features are less relevant than others in communication. Some features are more hidden and not easily communicated or transmitted. Something like religion is difficult to transmit and Axelrod just assumes all features are easily transmitted. The other important characteristic not incorporated in the model is the negligence of the dependence of cultural features.

The next section is background information about the theory for language and the mechanism of language transmission from adult to child. The theory of language is Stochastic Optimality Theory [Boersma and Hayes, 2001] where the mechanism for language transmission is the Maximal Gradual Learning Algorithm, also [Boersma and Hayes, 2001]

Chapter 3

An Overview of Stochastic Optimality Theory

In this project there is need for an adequate or biological plausible representation of language. It would be unconvincing to just invent a language. There needs to be a language model, which is firmly grounded in the theory of human language. This language cannot be as complex as human language, but can be an abstraction of it.

The representation of language used in this project, is inspired by *Stochastic Optimality Theory* (SOT) developed by Paul Boersma; see Boersma [1997] and Boersma and Hayes [2001]. Stochastic Optimality Theory is an extension of the Optimality Theory (OT) developed by Prince and Smolensky [1993]. Because SOT is an extension of OT it is only logical to first start the discussion about OT.

3.1 Optimality Theory

As Prince and Smolensky [1997] point out that there are many differences between languages. But languages have also many structural items in common. They have tried to find an architecture, which is capable of modeling the differences and also the commonalities, and have consequently developed their Optimality Theory.

Prince and Smolensky [1997] have observed that grammars have multiple constraints on the well-formedness of linguistic structures. These constraints can be in conflict. A very simple example where this occurs is where there is a conflict between the basic structural word order subject-verb-object constraint and the constraint that puts the question word first. This example is from Prince and Smolensky [1997], but than translated into a Dutch version. The basic structural word order can be observed in the sentence *Piet ziet een haas* (translated: Piet sees a hare). When this sentence is put into a question form, *Wat ziet Piet?* (translated: What sees Piet?), the object is first. So there are two sentences in which two constraints are in conflict. In the first sentence the subject-verb-object constraint is stronger and in the other the constraint for putting the question word first.

Conflicts between forms can also occur between levels of language. Prince and Smolensky [1997] give a nice example in English. The English past-tense form of "slip", "slipped", is pronounced as *slipt*. A general phonological constraint on devoicing in final consonant sequences favors the harmonizing voicing pronunciation pt over pd. This conflicts with the well-formedness constraint that the past-tense's basic form -d is given. English solves this conflict by ranking the phonological constraint higher than the faithfulness constraint. Other languages might resolve this differently.

Prince and Smolensky [1997] have, from multiple examples where there are multiple conflicting constraints, extracted the hypothesis that these constraints are strictly ranked. This means that the violation of a higher ranked constraint is always a more serious violation than the worst possible violation of all the lower ranked constraints.

Prince and Smolensky have come up with a mechanism to determine, based on this strict constraint ranking, which form is optimal. The first thing to know is that there is an *underlying form*, out of which the *candidate forms* are generated. This underlying form is universal for every language. The multiple candidate forms, are generated by a *Generator* or officially GEN and are also universal for every language. These candidate forms are evaluated by the EVAL. For the specific underlying form which is being evaluated, several constraints are of concern. Although every form that is generated violates some constraint(s), an optimal form can be determined by finding, done by EVAL, the single form that has minimally violated the ranked constraints. Which candidate form becomes optimal, depends on the order of the constraints. This order differs for every language.

Optimality Theory holds that there is a universal set of *well-formedness* constraints [Prince and Smolensky, 1997]. Optimality theory also entails that the forms that are competing for the optimal form are universal. These two things are what grammars of different languages have in common. The strict ranking of the constraints varies and these variations accounts for different grammars of different languages. It is common that *EVAL*, with a different ranking of constraints, evaluates a different optimal form out of the universal set of forms.

Table 3.1: Example of a constraint ranking in Optimality Theory

/I language want learn to/			A1	A2
a		*language to I learn language*	*!	
b	☞	*I want to learn language*		*
c		*to want learn language I*	*!	**
d		*I language want to learn*		**!*

/I language want learn to/			A2	A1
a	☞	*I language want to learn*		*
b		*I want to learn language*	*!	
c		*to want learn language I*	*!*	*
d		*language to I learn language*	*!**	

To show how the GEN and the EVAL work look at table 3.1 , where two tableaux of conflicting constraints are shown. This example is made up and the constraints, the candidate forms and the chosen optimal form have no meaning; it is merely to explain the mechanism and its consequences. Although the constraints have no meaning, they have similar names as the constraints in the model used in this project. Each tableau represents a different constraint ordering and consequently a different grammar. In the left top corner the underlying form */I language want learn to/* is given. This underlying form is the input for the generator *GEN*. The *GEN* produces several candidate forms out of the underlying form. Below */I language want learn to/* some of these forms are given. From the second column onward in the first row the constraints are represented, which are ranked from left to right. Next to every candidate form, the violations of this candidate form of the constraints are given. An asterisk means the candidate form violates the constraint and an exclamation point means that this specific violation is fatal, consequently signalling that the candidate is a suboptimal form. The hand points to the optimal form in the current constraint ranking. In this specific case the constraint A1 and the constraint A2 are used.

The two tableaux represent different grammars, because the strict constraint ranking is different. In tableau **A**, constraint A1 is ranked higher. */I want to learn language/* is signalled as the optimal form here, because A1 is not violated. But */I language want to learn/* also does not violate A1. How to determine which of these two candidate forms is optimal? */I language want to learn/* and */I want to learn language/* both violate A2, but */I language want to learn/* violates it more because there are three *s. The choice of violations is arbitrary, so please refrain yourself from searching any meaning behind it. */I want to learn language/* violates A2 only once. In tableau **B** A2 is ranked higher. In this grammar only one candidate form, *I language want to learn*, does not violate the A2. Because this is the only form, it is easily evaluated as the optimal form. This example illustrates that a simple exchange of the ranking of constraints can account for different optimal forms and consequently for different grammars. *I language want to learn* could perfectly be a well-formed sentence in another language

Prince and Smolensky claim that OT is a good theory for the phonological and grammatical properties of languages. Simple strict domination hiëarchy of well-formedness constraints can account for many complex grammatical structures and different rankings of the constraint set can account for many different linguistic patterns. They claim they have accounted for the basic prerequisites for a theory of language, which must be built out of simple parts. With different combinations of these parts or constraints, different kinds of linguistic patterns can be formed. The question: *What do the grammars of different languages have in common, and how may they differ?* is hereby directly answered.

3.2 Stochastic Optimality Theory

The standard OT theory has the assumption that the constraints are strictly ranked [Boersma, 1997]. But there are several cases where multiple options are possible, entailing optionality. Standard OT cannot cope with this optionality in language because of this strict ranking. In SOT the constraints are ranked along a continuous scale. The ranking is not fixed, but at evaluation time a normally distributed noise value (stochastic) is added to the standard ranking, causing variation. Close constraints will therefore be able to be switched causing different outputs, hence optionality.

SOT is applied to very specific examples Boersma [1997], but aims at being able to learn complete languages. SOT is applicable at every level of language and aims to account for an explanation for the learning of variation, optionality and probability. In this project this mechanism of representing and learning language is used. Boersma [1997] mentions a typical example

Table 3.2: Example of a case where optionality will occur. This is the case where [anpa] will win over [ampa] [Boersma, 1997]

|an+pa|	B1	B2
☞ [anpa] /anpa/		*
[ampa] /ampa/	*!	

of optionality in speech production. When a word ends with |an| and is directly followed by a |pa|, this may be pronounced as either [anpa] or [ampa] by an individual. This typical example is

Table 3.3: Example of a case where optionality will occur. This is the case where [ampa] will win over [anpa] [Boersma, 1997] ..

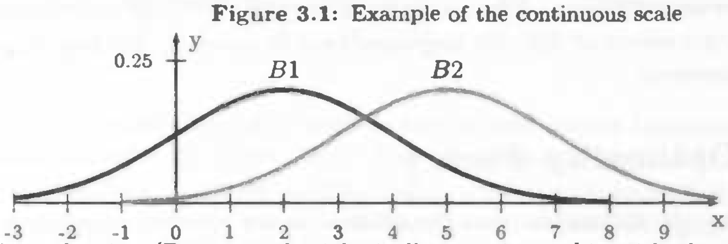
|an+pa|	B2	B1
[anpa] /anpa/	*!	
☞ [ampa] /ampa/		*

hard for Optimality Theory [Prince and Smolensky, 1993] to solve, because of the strict constraint ranking. There are two constraints where presently the meaning of them does not matter. For convenience they are again named in the way the constraints are named in the model in this thesis: B1 and B2. The candidate [anpa] will win if the ranking is B1 >> B2. This is illustrated in 3.2. But the candidate [ampa] will win when the ranking is B2 >> B1, see figure 3.3. Because the strict ranking of OT has a problem Boersma [1997] suggests that ranking should be considered in a probabilistic manner.

3.2.1 Stochastic Disharmony

Boersma [1997] solves the problem of optionality by placing the constraints along a *continuous scale*. A random element is added to every constraint's ranking value during evaluation or *EVAL*. In figure 3.1 this phenomenon is displayed. B1 is placed on real value 2 and B2 at 5. With

Figure 3.1: Example of the continuous scale



stochastic evaluation (Boersma physiologically associates this with the noise in the amount of locally available neurotransmitter), the order of the constraints is determined by the ranking at evaluation time, *disharmonies*, of the constraints which are evaluated as follows [Boersma, 1997, page 45]:

$$disharmony = ranking + rankingSpreading \cdot \mathbf{z} \qquad (3.1)$$

z is a Gaussian random variable with mean 0 and standard deviation 1. An example of ten evaluations of —an+pa— with a *rankingSpreading* of 2 is given in figure 3.4. In most of the

Table 3.4: 10 trials with the constraint ranking in 3.4
. Based on the example given in [Boersma, 1997, page 45]

trial	1	2	3	4	5	6	7	8	9	10
B1 disharmony	3.5	4.2	3.2	2.1	5.9	5.9	5.7	6.8	8.4	7.3
B2 disharmony	3.8	1.3	3.7	2.2	1.9	1.8	1.2	3.3	1.1	1.7
outcome	np	mp	np	np	mp	mp	mp	mp	mp	mp

evaluations B1 comes out higher but sometimes also B2 wins. In figure 3.4, [ampa] would be pronounced 7 and [anpa] 3 times. The calculation to get the chance that one disharmony becomes higher than the other is [Boersma, 1997, page 45]:

$$P(disharmony_1 > disharmony_2) = \frac{1}{2} \cdot \left(1 - erf\left(\frac{1}{2}\sqrt{2} \cdot \frac{r1 - r2}{rankingSpreading \cdot \sqrt{2}}\right)\right) \quad (3.2)$$

Boersma claims that when the ranking difference is below ten it is possible to speak of *optionality*, otherwise of *obligation*.

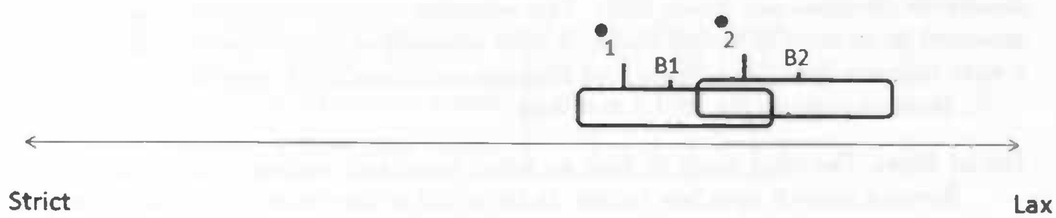
The agents in the simulation will possess a language described by a few constraints, ranked along this continuous scale. Every child-agent will have a critical period in which it is able to learn language. This learning will be done by the Maximal Gradual Learning Algorithm (MGLA), also developed by Boersma [1997].

3.2.2 Maximal Gradual Learning Algorithm (MGLA)

Boersma [1997] shows some very simple and robust gradual language learning algorithms. The Maximal Gradual Learning Algorithm which Boersma thinks is best is discussed here.

Before explaining the learning of the algorithm it is important to know the representation of the constraints and the details of free variation. As said earlier there is the *continuous ranking scale*. The constraints all have a ranking value along this continuous scale. Then there is the *stochastic candidate evaluation*, where a stochastic element is added to each constraint. When constraints are close in real values as in figure 3.1 things become interesting but another, more systematic example is given in 3.2 and 3.3. This example is highly similar to the example given in Boersma and Hayes [2001, page 3] There are two constraints B1 and B2 which overlap. Most

Figure 3.2: Constraints B1 and B2 overlap. The dots illustrate the outputs of constraint B1 and B2. Here B1 is ranked higher than B2 [Boersma and Hayes, 2001, page 4]



of the time, B1 will be ranked higher, but as in 3.3 an occasional switching occurs. This is *free variation*, were multiple outputs are produced out of a single underlying form.

The constraint ranges are represented as *probability distributions*. In this case Boersma uses a *normal* or *Gaussian distribution*. Boersma makes several assumptions for the Gradual Learning Algorithm. The first assumption is that the selection points for natural language constraints are distributed normally. The second assumption is that every constraint has the same standard deviation. This entails that every constraint depends only on its ranking values. The term *evaluation noise* is used to describe the standard deviation.

The constraints are displayed in figure 3.4 with an example of overlapping ranking distributions. The standard deviation is 2.0 and the value of the constraints are B1 (5) and B2 (9). With equation 3.2 the probability of a switching of constraints can be calculated, which are 7.9% for the case of B2 >> B1 and 92.1% for B1 >> B2.

Figure 3.3: Same ranking as in 3.2 but in this the rare case where B2 is ranked higher than B1 [Boersma and Hayes, 2001, page 4]

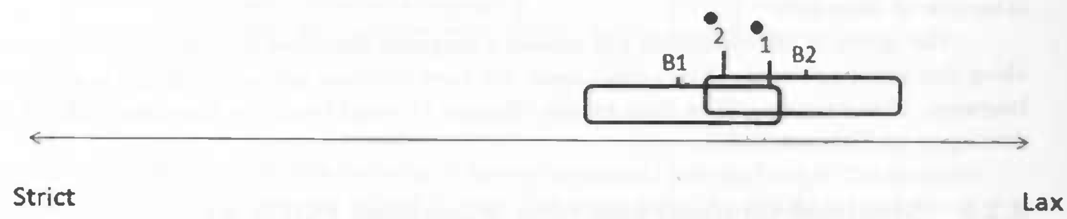
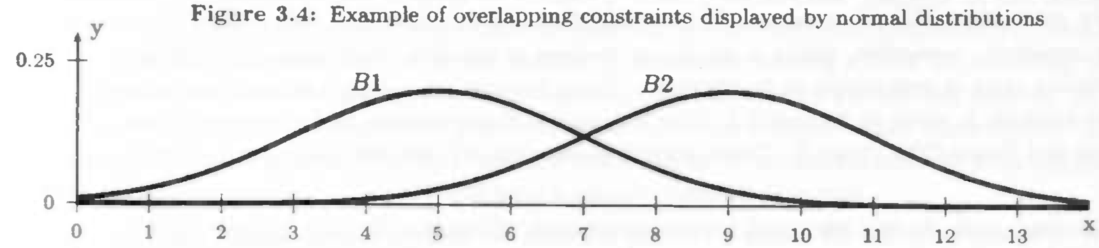


Figure 3.4: Example of overlapping constraints displayed by normal distributions



Learning

The Gradual Learning Algorithm aims to have a child learn a language by slowly adjusting the constraints [Boersma and Hayes, 2001]. This adjusting is to be done by learning from utterances presented by an adult or several adults. A short summary of the mechanism is given here but for a more complete description please read Boersma and Hayes [2001, page 6-8]

Boersma presents the MGLA in 4 steps:

Initial State The child needs to have an initial constraint ranking to represent its grammar. Boersma leaves it open how to this. In the model in this thesis, all the adults start with the same constraint ranking values, which are randomly chosen and the child starts with all the constraints ranked 0.

Step1: A datum The child is presented with an adult utterance or as Boersma calls it, a learning datum. Boersma does not say how this learning datum is generated, but here it is assumed that it is generated the same way as the child's grammar generates for the specific underlying form. This generation of an utterance is subject to *imperfect performance*

Step2: Generation The output of the child's grammar output is generated from the /underlying form/ by *stochastic candidate evaluation* for every single constraint. This means a noise value is added to every constraint to obtain the so-called selection point. Now that every constraint has a selection point, they are ordered from high to low. After this the constraints are evaluated just as in the normal Optimality Theory and an optimal form arises.

Step3: Comparison The child assumes the learning datum to be correct and compares its own output with the learning datum. If the output is different from the learning datum, the child will learn. In table 3.5 an example of difference is given which is highly inspired upon the example given in Boersma and Hayes [2001, page 7]. In the first column, the 2 candidate forms are presented, the learning datum and the learnerś output respectively. As one can see, the child has B1 ranked too high. Because several constraint violations do not influence the outcome, these are canceled by a mechanism Boersma calls *mark cancelation*, see table 3.6. After this the adjustment of the constraints of the child can take place.

Step4: Adjustment The adjustment takes place by decreasing the uncanceled violated constraints of the learning datum and increasing the uncanceled violated constraints of the child's output, see figure 3.7. The amount of adjustment is determined by a small value, which is called *plasticity*. This small adjustment is a discrete value, which means the perfect ranking is never learned. This means there is also a certain amount of *imperfect learning*, determined by the small value. The small arrows illustrate in which direction the constraints are adjusted. Usually the *plasticity* value will be small and a key hypotheses of the Gradual Learning Algorithm is:

- Small adjustments of ranking values are made.

But, actually there must be some relation between the plasticity value, the standard deviation of the normal distribution and the number of learning steps. Logically, the plasticity must not be so high that a constraint can move more than one standard deviation, else it overshoots. Plasticity must not be too low, because learning will be to slow.

Final state This process is repeated as more learning data is presented. If the output for a certain underlying form stays the same, the child-learner will eventually learn the right constraint ranking, in the sense that the rankings are so far apart that the probability of another candidate becomes eventually zero.

Table 3.5: The learner's output is different from the adult's output

/underlying from/	*B*1	*B*2	*B*3	*B*4
√ Candidate 1 (learning datum)	*!	* *	*	
☞ Candidate 2 (learnerś output)		*	*	*

Table 3.6: Mark cancelation

/underlying from/	*B*1	*B*2	*B*3	*B*4
√ Candidate 1 (learning datum)	*!	* $\cancel{*}$	$\cancel{*}$	
☞ Candidate 2 (learnerś output)		$\cancel{*}$	$\cancel{*}$	*

3.2.3 How does free Variation work

The question is now: How can SOT account for free variation as in the case of /anpa/ and /ampa/, where multiple output forms are possible out of a single underlying form? The child's grammar will

Table 3.7: The constraint demotion and promotion

/underlying from/	*B*1	*B*2	*B*3	*B*4
√ Candidate 1 (learning datum)	* →	* →		
☞ Candidate 2 (learner's output)				← *

mimic the relative frequency of the free variants in the environment. Lets go back to the example of /anpa/ and /ampa/. *GESTURE and *REPLACE are the constraints. As a child is represented /anpa/ 64% and /ampa/ 36% of the time, and the settings are as in 3.2, the ranking difference will be approximately 1 between *REPLACE and *GESTURE, where the closeness depends on the absolute frequency of representations. The child-agent will have a finite number of encounters with adult-speakers who present an utterance. This utterance, at evaluation time, is based on the adultś language, which consists of a ranking of their constraints. Because of the stochastic elements added to the several constraints, the utterances vary every time (imperfect performance). At each encounter the child-agent will update its constraint ranking a little bit according to the difference between the adult-speaker's utterance and its one temporary constraint ranking. At last, the child-agent's constraint ranking is fixed when it reaches adulthood and consequently cannot learn anymore. Because of the variation in the utterances it encounters from different adults and of the finite encounters, the child will have a slightly different language.

SOT aims to explain optionality and robustness in language, but in this thesis it will be used for simulating the discrepancy between linguistic competence and performance. This discontinuity in language transmission reflects the idea posed by Meillet (Meillet 1926), (see 2.6.1), that imperfect performance and learning causes language variation. This is very important to note, because this implies the source of variation of language, and has a clear connection with the nature of language. The hypothesis here is that, in time, language will change.

A note has to be made on the question of free variation. Stochastic OT can account for free variation but variation is not always free because extra-linguistic factors (class, age and sex) can determine the one or the other form. Example is the 'cool' use of *cah* in stead of *car*.

Chapter 4

Behavior of the System

In this project two models have been developed. First, an *Individual Language Model* (ILM) has been developed to investigate the transmission between a single adult and a single child agent, the evolution of language over multiple generations of single adult and child agent transmissions and the applicability of Stochastic Optimality Theory. The representation and mechanism of language transmission developed in the ILM are fully implemented into the Spatial Population Model (SPM). The Spatial Population Model is a complex dynamic system (see chapter 2.5) where global linguistic behavior emerges from the local interactions between multiple child and adult agents. Consequently, there is no *central authority*. There is a population of agents spread spatially over a 2D world where multiple children learn language from their parent and possibly other agents and where the development of linguistic diversity can be simulated.

The SPM and the ILM are computer simulations which involves making a simple set of assumptions about how the process works, then implementing this in an idealized set of agents. The simple set of assumptions in the SPM and ILM are:

- **The world** where the agents live is simple. It is a 2D grid with discrete sites which can each be occupied by multiple agents.
- **Time** changes the world. Time is simulated in discrete steps. At every step, the world is updated. Updating means an iteration over all the agents, which will execute their actions.
- **The agents** have the following properties:
  - The agents have age, and over time they die.
  - The agents all go through a period of childhood, where the only differences in comparison with adults are that they are able to learn language and that they move with their parents
  - The ability to move over the 2D grid.
  - The ability to reproduce in the reproductive period.
  - They possess language. Child agents learn language from their parent and the choice can be made to let children learn from others from a certain age.
- **The language** is a theoretical based abstraction of real language.
- **The language transmission** from adults to children is done with a certain discontinuity (imperfect learning and performance), resulting in a slight variation in the language of the children in comparison with the adults they learned from.

- The **parameters** are variables which can be chosen by the researcher to investigate what he desires. These constrain the world, the agents, the language and the language transmission, but ultimately the whole simulation run.

So, the SPM consists of all these aspects and the ILM only of the language and language transmission part.

The mechanisms of language evolution; discontinuity of language transmission, the spatial organization and mobility are all incorporated in this initial model. With several parameters the first mechanism can be tuned in order to cause a higher or lower discontinuity. Mobility can be switched off entirely, but can also be tuned higher or lower. The influence of the spatial organization can also be tuned lower or higher, by varying several parameters. The high flexibility in adjusting the influence of these three mechanisms makes it possible to investigate a large space of possibilities.

A few questions arise which are important to consider:

1. How will the agents be distributed over space initially?
2. Will they start with few and at a single location, supporting the monogenesis theory, or will they start at different places (polygenesis)?
3. Will the agents move across space and when they move how will they move?
4. How is the language of an agent represented?
5. How are languages learned by the children?
6. How to decide whether two languages are two different languages or just dialects?

The last three questions are answered in the following chapter and the rest in the chapter describing the SPM. Because first the ILM has been developed the first section will be about this model. In this section the representation of the language, the mechanism of language transmission and the Iterated Learning Model (IL) [Kirby, 2001] are explained. The mechanism to simulate generations following up generations is based on the IL. Thereafter, the SPM is treated.

4.1 Individual Language Model (ILM)

4.1.1 Representation of Language

The representation of human language that is used in the ILM and then later also in the SPM, is inspired by Stochastic Optimality Theory (SOT) developed by Paul Boersma; see Boersma [1997] and section 3.2. This representation should be sufficient to demonstrate the evolution of diversity of language. As said in 3.1 Prince and Smolensky [1997], languages *differ* and languages have properties in *common*. What languages have in *common* is a *universal* set of constraints. In this model a universally fixed set of constraints is used. These constraints are ranked along a continuous scale, see figure 4.1. Differences in constraint ranking but not in constraint order account for different grammars within the same language, and different constraint orders for different languages. That different constraint orders, or discrete differences, are more important is also supported by Nettle [1999] who claims that most linguistic variables are not *continuous* but *discrete*.

The representation and mechanism of language transmission in the ILM are nonetheless a simplification of Stochastic Optimality Theory. Several characteristics of SOT are not used in this model. Optimality Theory produces several forms out of the underlying form, see figure 3.1. A generator, *GEN* generates these forms. Next, the constraints are evaluated by *EVAL*, by marking the constraint violations. At last, the optimal form is found. This process of generating forms and evaluating these is not used here. The only characteristic that remains in this model is the ranking of several constraints along the language scale and the child needs to learn this constraint ranking. So the child only learns the constraint ranking. Normally, a child has a certain conviction about a certain optimal form. This conviction is evaluated from the constraint ranking in the child. This conviction is the optimal form. Therefore the hypothesis here is that the constraint ranking is essentially the basic description of a language, from which the optimal form is found. Because the essential part of the Optimality Theory is used, this is a plausible abstraction. The language of an agent therefore consists of a ranking of the universal set of constraints. Normally, these constraints have a meaning, but here this also is left out. The universal constraint set also has many more constraints than in this model.

The following example shows how the SOT is represented in representation of language in the ILM. In figure 4.1, the example of the case with */I language want learn to/* is shown again, which is explained fully in section 3.1. What is important here is that there are the two constraints: A1 and A2. Only the two constraints A1 and A2 would be represented in the ILM. The several candidate forms are not present in the ILM. Consequently, there is not an evaluation of these forms along the constraints. The constraints do also not have a meaning, as it normally would have. One can imagine that there are many constraints that make up all options in a language. In this language model the amount of constraints is small, approximately 2 to 20.

Table 4.1: Example of a constraint ranking in Optimality Theory

/I language want learn to/		A1	A2
a	*language to I learn language*	*!	
b	☞ *I want to learn language*		*
c	*to want learn language I*	*!	**
d	*I language want to learn*		**!*

How a single constraint is represented is shown in 4.1. The constraint is drawn as a normal distribution. Every constraint has a different color and a label attached to it. The blue constraint is labeled A1 for example. The agent has the constraint ranked at where the top of the normal distribution lies. In the figure some normal distributions of constraints can overlap. Essentially figure 4.1 is similar to figure 3.4.

The constraints are, as in SOT, placed along a continuous line. The range of the *continuous scale* is fixed and limited between 0 and a given LANGUAGESCALE value. The constraints therefore have a limited value range. This value range can be adjusted by adjusting the LANGUAGESCALE parameter. A larger continuous scale indicates a lower amount of noise and a smaller one a higher amount of noise. The last is because the normal distributions of the constraints lie closer together, consequently have a larger chance of being exchanged. Passive constraints are placed on zero. A bigger scale decreases the chance of producing an innovation or variation, because the constraints are generally further apart from 0.

It is crucial to know that the constraints in the language are for each adult divided into passive and active constraints. The passive constraints all lie at zero indicated by the text *passive*, where the active constraints lie above zero and are not drawn in the figure. A language is defined by the active constraints. Consequently, the language of the adult given in figure 4.1, is represented by the active constraints [A8=0.4, A1=10.6, A4=14.1, A6=31.8, A2=40.3, A3=40.8].

Figure 4.1: The initial adult language with values: [A8=0.4, A1=10.6, A4=14.1, A6=31.8, A2=39, A3=42]

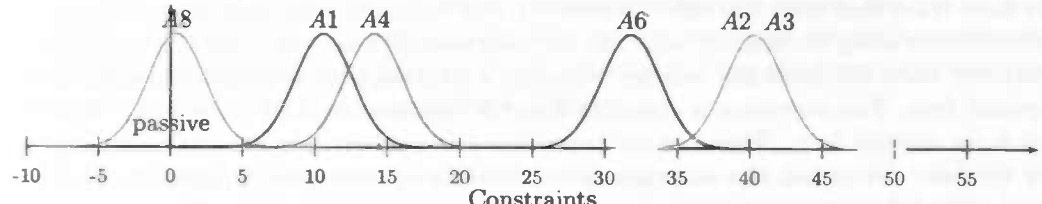
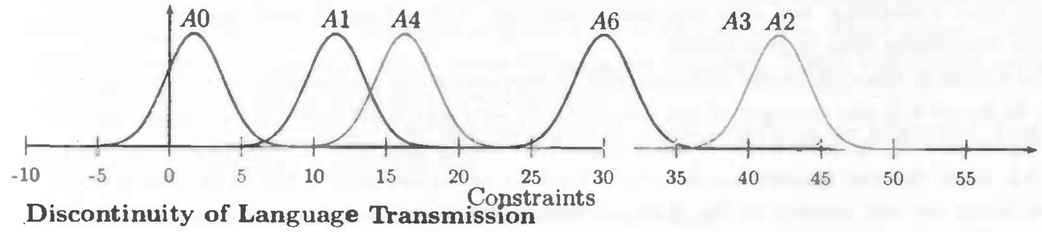


Figure 4.2: An utterance by the adult with values: [A0=1.7, A1=11.5, A4=16.3, A6=30, A3=39.9, A2=42.1]



Discontinuity of Language Transmission

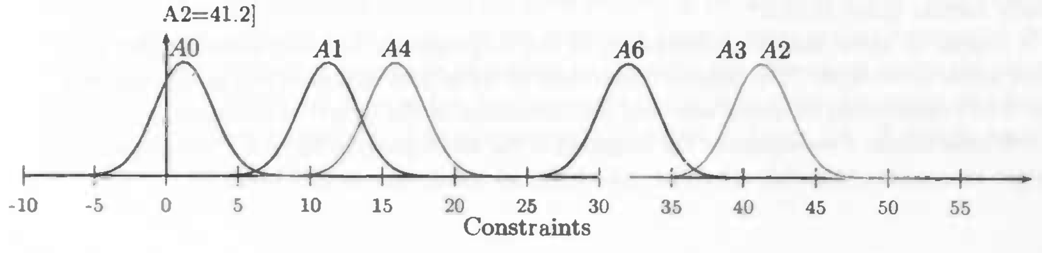
The variation in the utterances that the adult produces means that there is a DLT. This is very important to note, because DLT implies a *source of variation of language*, and has a clear connection with the nature of language. In section 2.6.1 two types of DLT are discussed. In this language model the variation in utterances corresponds to *imperfect performance*, where there is a discrepancy between linguistic competence and performance or adult language and its utterances and the discrete learn rate PLASTICITY to *imperfect learning*. This discrepancy between linguistic competence and performance is affected by the size of the LANGUAGESCALE. Although SOT aims to explain optionality and robustness in language, it can also be used to model DLT. The hypothesis here is that because of the DLT, in time, language changes.

4.1.2 Mechanism of Language Learning

The mechanism by which children learn the language from adults is based on the Maximal Gradual Learning Algorithm (MGLA), and is summarized in 3.2.2. Just as the representation of language is a simplification of the representation in Optimality Theory, the learning mechanism in this model is a simplification of the MGLA. In the representation of language there are no different candidate forms of an underlying form, which are evaluated according to the constraint ranking. Consequently, the child and the adult do not perform the mark cancelation step in the MGLA. Step 2, the generation, is also left out, because the child does not produce an output. What remains in this language model are the stochastic production of constraint rankings in the form of an utterance and the learning of the constraint ranking by a child.

As said in section 2.6.1, the model in this thesis would be achieved by the two ways of

Figure 4.3: The language learned by the child with values: [A0=1.2, A1=11.2, A4=15.9, A6=32, A3=40, A2=41.2]



Discontinuity of Language Transmission (DLT): imperfect performance and imperfect learning. *Imperfect performance* expresses itself through the stochastic production of the adult's utterance, and the *imperfect learning* through discrete learning steps of the child.

First of all, every child agent has a critical period in which it is able to learn language. When the child reaches adulthood it does not learn anymore. Due to imperfect performance and imperfect learning the child learns is a minor random variant of the grammar taught by the parent and perhaps by others, similar to other models from Nettle [1999] and Steels and McIntyre [1997].

During this critical period the child-agent has a finite number of encounters or learning moments with adult-speakers. An encounter consists of two general steps:

1. The adult produces an utterance from its language.
2. The child learns from this utterance.

An utterance, at evaluation time, is based on the adults' language, which consists of a subset of the active constraints. Because of the stochastic elements added to the several constraints, the utterances vary every time (imperfect performance). Every constraint ranking will deviate by a value drawn, another for every constraint, from the *normal distribution*. The normal distribution has a mean μ equal to the constraint ranking. The standard deviation σ is constant at 2. This means the normal distribution for every constraint is the same through every experiment. The adult evaluates every constraint by calculating the *disharmony* or ranking at evaluation time of these constraints by equation 4.1. For an explanation of this equation see section 3.2.1.

$$disharmony = ranking + rankingSpreading \cdot z \tag{4.1}$$

Next, the adult takes the fixed number, determined by the parameter NUMACTIVECONSTRAINTS, of highest constraints as its active constraints. From these temporarily active constraints a fixed number determined by the parameter NUMUTTERANCECONSTRAINTS, is taken as the utterance. The collection of these constraints form the utterance. This way the child receives an utterance containing constraint values which vary slightly from the constraint ranking of the adult (imperfect performance). Naturally, there is a slight chance that every stochastic element for the constraint have a value of zero, causing no variation in the utterance. In figure 4.2 an utterance is presented by the adult with the language in figure 4.1. In this case the subset is a total subset of the active constraints: every constraint in the adult language, or active constraints, is present in the utterance. In the figure the overlaying normal distributions are also shown, but in the transition the utterance does not contain this information. The utterance clearly deviates from the ranking of the constraints of the adult language. A clear difference is the constraints A3 and A2 which are exchanged in the utterance, which are exchanged in the utterance. The probability an exchange takes place between two constraints given the constraint rankings is calculated by equation 4.2.

$$P(disharmony_1 > disharmony_2) = \frac{1}{2} \cdot \left(1 - erf\left(\frac{1}{2}\sqrt{2} \cdot \frac{r1 - r2}{rankingSpreading \cdot \sqrt{2}}\right)\right) \tag{4.2}$$

At each encounter the child-agent updates its constraint ranking a little bit according to the difference between the adult-speaker's utterance and temporary constraint ranking. If the constraint is ranked higher than the constraint of the adult, the constraint ranking is reduced or else it is increased. This is essentially equal to the operation done by the MGLA. The amount of adjustment (or the amount of imperfect learning) of a constraint is determined by the PLASTICITY or learning rate, which can be adjusted. If p is high the learning will be faster, but will be less accurate. When a constraint ranking of the child is near the constraint ranking of the adult at time t, with a high p the constraint ranking in $t+1$ has a greater probability of overshooting than when p is lower.

The number of encounters between a child and adult(s) is limited and can be chosen for with the parameter LEARNSTEPS. Finally, after the last encounter, the child agent's constraint ranking is fixed, because of reaching adulthood. Whether this new grammar is a different language depends on if the order of the constraints is different.

4.1.3 The Development of New Languages

The development of new languages can occur in two ways. The first is in the form of *simple variation*: A different order of the active constraints in the child's language in comparison with the language of the adult. The second is in the form of an innovation, where a new constraint is present in the child language, which is not in the adult language. These innovations correspond to for example new words or sounds in a language.

In the model there is need for the development of innovations in the languages. Essentially there are two options to accomplish this:

1. The emergence of a new language feature or constraint.
2. Have a fixed set of language features, with some features active and some passive. The active features describe the language and the passive features are the non-participating ones. A new feature emerges when a passive feature becomes active.

In this project the second option is chosen. Every agent has a language containing the same set of constraints, but the distribution of or the ranking values of these on the continuous scale are different with different grammars. The passive and active features are the passive and active constraints. At the initial stage of the simulation, every agent has the same language in that the ranking values of all the constraints are exactly the same. This language consists of all the constraints. Some fixed amount of constraints are active and have a real value above zero. The other constraints, for which the number is also fixed, are passive and are zero. These passive constraints are analogous to the hidden features which are not represented in the specific language. In figure 4.1 an example of an adult's language is shown. Six constraints are active and the constraints at zero, which are invisible, are the passive ones.

During the learning fase, the adult agent presents its language to the child-agent by producing utterances. In the example, the language consists of six active constraints. The child agent's constraints are initially all ranked zero. Every child has to start with the same empty language, since the languages between newborns of different language regions do not differ and because in this thesis the conviction is that language is non-innate but socially and culturally transmitted. Then learning starts and the two types of variation can occur.

Simple Variation

The first mentioned type of variation is simple variation. In the figures 4.1, 4.2 and 4.3, respectively the language of the adult, an utterance produced by this adult and the learned language of the child are shown. In this example simple variation occurs. In the adult language, active constraints *A2* and *A3* are in the order [A2, A3]. In the utterance these constraints are exchanged [A3, A2]. When the majority of the utterances presented to the child are utterances where the constraints *A2* and *A3* are exchanged, than in the final grammar of the child, when becoming an adult, these constraints are also exchanged. This phenomenon is shown in figure 4.3. A simple variation in the language has occurred.

Innovation

When a single active constraint lies close to zero, thus near the set of passive constraints, there is a reasonable chance the adult swaps these constraints in the output. If this happens, a passive constraint gets actively involved in the learning process. The adult temporarily judges the passive constraint as an active constraint and the active constraint as a passive constraint. In figure 4.2 an example of this is shown. The gray constraint *A0* is new in the utterance and the adult has judged this passive constraint as an active constraint. The blue constraint *A8* is judged as a passive constraint in the utterance, hence can not be seen. When the majority of the utterances presented to the child are utterances where the constraint *A8* is a passive and *A0* is an active constraint, the learner adopts this as an active constraint in its adult fase. An innovation has evolved in the language.

4.1.4 Other Issues

It has to be tested how many iterations there are needed to accurately learn the language. The number of iterations depend on the PLASTICITY value, the size of the continuous scale in which the constraints lie and the percentage of language which gets represented in every utterance. These are experimental issues. The characteristics of the normal distribution are fixed.

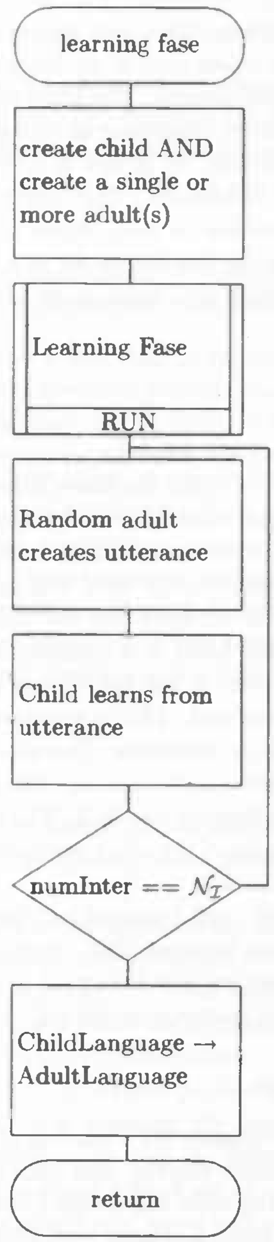
A choice has to made here between presenting the total language or a subset to the child. A subset would be more natural since children do not get to see the whole language at once. Instead they only get to see a small part. On the other hand, it is computationally costly to only present part of the language to them at every encounter. It will take longer to learn the language as well as in the case where the total language is presented. This is something to experiment on.

Another issue is the speed of innovation in languages. This also depends on the PLASTICITY value and the size of the domain in which the constraints lie. The PLASTICITY determines the degree of imperfect learning and therefore influences the DLT. The size of the domain indirectly determines the amount of noise in the utterances of the adult and therefore the degree of imperfect performance by the adult and thus also the DLT.

In general there has to be a qualitatively well balanced individual language learning model. With this individual model initial testing can be done. But, especially the plasticity-value, the percentage of language presented in an utterance and the speed of innovation are varied in the SPM in order to experiment on the effects on language evolution.

4.1.5 The Total Learning Phase

The total learning fase in the Individual Stochastic OT Learning of the child is a fixed number of interactions *learnSteps* $\mathcal{N}_{\mathcal{I}}$ with one or more adult(s) and corresponds to a single *run*. The idea is shown in the diagram below. When the child has finished the learning fase it becomes an adult and the language is transformed to an adult language, by taking the fixed number of highest constraints as active constraints.



4.1.6 The Iterated Learning Model

The Individual Learning Model is also used to experiment on the development of language over generations. For this an *Iterated Learning Model* (IL) is used. The theory of IL is from Kirby [2001]. The essence of the IL lies in just the development of a certain feature, which could be a language, over generations.

In this model an adult teaches a child its grammar. When the child has done the decided number of interactions it becomes the adult and teaches a new child its language. This process is repeated until a stop is forced. In this way the change over generations can be tested with different parameter settings. The general hypothesis is:

> Over the generations the grammars differentiate more and more from the initial language of the first adult.

4.2 The Spatial Population Model

Now that the representation and transmission of language are explained it is time to explain the Spatial Population Model where the representation and transmission of language are incorporated. The SPM is programmed with Java in Eclipse, with the aid of the java- based Recursive Porus Agent Simulation Toolkit (RePAST), developed by the Repast Organization for Architecture and Development[Collier et al., 2003]. The *sugar scape* example of a Repast implementation, which can be found on the site of Repast [Collier et al., 2003], is the starting point. The sugar scape has some similarities with this model. These are that the model consists of agents moving on a 2D grid, which live for a limited time. This sugar model has been adjusted and expanded to the model's wishes.

4.2.1 Parameters

At the start of this section all the parameters in the model are summarized in table 4.2, for the readers convenience. Later in this section it is then possible to refer to the parameters. In the table, the default values for every parameter are also shown. These values are the standard values by which the experiments with the SPM start.

4.2.2 Choices

The model is very flexible, and many parameters can be adjusted. This makes it a model with a high complexity, but gives a lot of freedom to the researcher to investigate a large variety of scenario's. Before the model starts to run the researcher chooses which parameter values to use. When the model has started it is not possible to change the parameter values in a single run.

Initialization

The initialization of the model is when the first step of the model is done, where several things happen. First of all, the model consists of a variable sized *grid* of *sites*, determined by the parameters GridXSize and GridYSize. Then, a variable number of *agents* are placed, determined by NumAgents, on the grid. *Multiple* agents are able to occupy a single site. Whether or not these agents are placed *randomly* across space can be determined by the parameter UniformSpaceDistribution. If this last parameter is checked, this entails a fixed number of agents are placed on every site. This fixed number can be determined by the parameter UniformAgentsPerCell.

Features of the Agents

The agents do not live eternally, but have a *variable age.* A MinDeathAge and a MaxDeathAge can be chosen. Every time a new agent is either born or initially placed on the grid, it receives a random age at which it dies, which lies between the MinDeathAge and MaxDeathAge. There is *no* gender, but whether all adult agents are able to *reproduce*, is decided by the parameter Replacement. If Replacement is chosen, agents which die are replaced by a new child-agent. There is no reproduction. If Replacement is not chosen, there is reproduction.

If Replacement is not chosen the agents have a certain *reproductive period* which can also be chosen. Initially, all the agents have an age randomly drawn, between the chosen parameters

MINREPRODUCEAGE and MAXREPRODUCEAGE. The minimum reproductive age is fixed at the ADULTAGE increased by 1, which also is a parameter. The choice to have the agents initially aged in the reproductive period, is based on the following arguments:

- The main argument is that if a random age is chosen there is a chance that only children or only adults above the maximum reproductive age are generated in the initialization fase. In the first case the children have no adults to learn language from, resulting in a non-language population. In the second case the 'oldies' are not able to reproduce, so that the population eventually dies out.
- Another argument is that with all the agents aged between reproductive ages, there is a large chance of survival of the population over time, because all the agents are able to reproduce.
- Also, the children that are born next start at age zero and have the whole length of childhood to learn the language. Otherwise it is possible that there are children of almost adult age. This is undesirable, because they do not have the full learning period.

Every agent is either a child agent or an adult agent. The child agent is able to learn language. Learning language stops at adulthood (The ADULTAGE can be chosen). The ADULTAGE and the LEARNSTEPS are mutually independent, they do not influence each other. The adult agent is not able to learn language, but is able to communicate its language to children.

The agents can be chosen to be MOBILE or not. If the agents are mobile, they move according to a simple *diffusion mechanism*. There simply is a small chance MOVECHANCE that the agent moves to another direct neighbor site. If the agent decides to move, then, randomly, one of the four spots is chosen.

Language

In 4.1 the representation and mechanism of language transmission is described. This model is largely copied into the SPM. Consequently, every adult agent has an adult language and every child a child language.

During the initialization fase, the agents, with ages between the *minimum* and *maximum* chosen reproductive age (parameters MINREPRODUCEAGE and MAXREPRODUCEAGE), all have exactly the same grammar and logically the same language. This initial language is randomly produced and is constrained by the A fixed set of constraints is produced, from which the active constraints and its rankings are randomly determined.

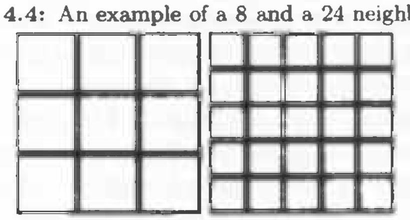
Throughout the simulation run certain properties of the language are fixed. The scale over which the constraints can lie is fixed by the parameter LANGUAGESCALE. This entails that the rankings of the constraints can only lie between 0 and a chosen upper bound LANGUAGESCALE. The language consists of a fixed set of constraints, determined by the parameter NUMCONSTRAINTS. Every constraint has a name and a value associated with it. The number of constraints can be chosen. The same is true for the number of active (parameter NUMACTIVECONSTRAINTS) and passive constraints (NUMCONSTRAINTS - NUMACTIVECONSTRAINTS), and the number of utterance constraints (parameter NUMUTTERANCECONSTRAINTS).

Then there is a fixed PLASTICITY or learnrate which can be chosen, and determines the speed and accuracy by which the language is learned. Finally, there is the fixed number of LEARNSTEPS the child gets, which corresponds to the number of interactions with adults in its childhood. It is important that the parameter settings are set so that the quality of the learning process is ensured. A number of experiments have been done to acquire a good representation of the results out of different parameter settings.

New Language Features In the SPM some features are added which are not represented in the individual learning model. These features are added because of the spatial component and the multiple agents present.

The first feature is the age at which the child is able to learn from non-parents (parameter LEARNFROMOTHERS). Before this age the child only learns from its parent. When the child agent grows older than the LEARNFROMOTHERS age, it has the chance to encounter other agents in its direct neighborhood. The size of this neighborhood is determined by the second feature and parameter LANGUAGEVISION. If for example the LANGUAGEVISION is set to 1, the neighborhood consists of the eight surrounding sites plus the site the child occupies, see the left figure in figure 4.4. If it is set on 2 there are 25 sites to choose from, displayed in the right figure. In case a child agent is able to learn from others, a random site is chosen. If there are adult agents on this site, the child agent randomly learns from one of these adult agents. At this age the child will randomly choose another adult in its direct surroundings to learn language from. When there is no possibility to learn from another agent, besides its parent, it learns from the parent. The values of the LEARNFROMOTHERS and LANGUAGEVISION parameters remain fixed throughout a simulation run.

Figure 4.4: An example of a 8 and a 24 neighborhood



When the child has an adult to communicate with, it learns the language exactly as in the individual language model, see section 4.1. The adult presents an utterance and next the child adjusts the constraint values of its own language by adding or subtracting the PLASTICITY value.

Development

During a simulation run there are developments of the model and the agents. A simulation run consists of simulation steps. At a single step all the agents are passed once in random order. Every child agent has learning moments with the parent or other adult agents and moves with its parent. Every adult agent moves or does not move (with parameter MOBILE), but takes its children with it wherever it goes. Every adult agent has a probability of teaching to children depending on several parameter settings (for example LEARNFROMOTHERS, LANGUAGEVISION and its situation (is it a parent or not).

When the possibility of reproduction is there, with the parameter REPLACEMENT unchecked, adult agents have a chance to reproduce. The growth of the population is controlled by controlling the chance of reproduction for every adult agent, depending on every agents' local surroundings.

Population Growth Control The logistic function is a common model of population growth. This function models the sigmoid-curve of growth. This function is based on two statements:

1. The population growth is based on the current population density
2. The population growth is based on the available resources As competition arises, the growth slows and at *maturity* the growth stops.

The equation used to determine the population growth is called the Verhulst Equation:

$$PopulationGrowth = \frac{dP}{dt} = rP(1 - \frac{P}{K}) \tag{4.3}$$

P is the current population density. r is the growth rate. K is the carrying capacity and is the population level that can be supported for an organism in a certain ecology. The carrying capacity is controlled by factors such as the quantity of resources and competition. The agents in this model do not have resources to compete for, such as sugar in the sugar scape model. But to implement a way in which population growth is controlled, and to simulate constraints posed by the environment this function is used.

In the SPM the population growth is controlled individually, by a *reproduction chance.* This seems logical, because an animal reproduces according to its local surroundings. So for every adult agent in the model which has an age in which it can reproduce, the local neighborhood is checked, and a chance on reproduction is calculated. Every adult agent which is able to reproduce calculates its chance on reproduction.

The parameter CARRYINGSCALE determines how local the population growth is controlled. If the local carrying scale is 1 the local 8 neighbors are considered. If the local carrying scale is 2 two rings around the site are considered, which are 24 neighbor sites. see figure 4.4.

The IDEALNUMAGENTSSPOT parameter determines how many agents are ideal for a site. The 4.3 equation is used to calculate the growth of the population in the neighborhood of the agent. K, the carrying capacity is determined by multiplying the *number of sites,* which is determined by the CARRYINGSCALE with the IDEALNUMAGENTSSPOT. The growth rate r is fixed at 0.5.

The personal birthrate per agent is calculated by dividing the determined *PopulationGrowth* for the neighborhood around the agent by the number of neighbors. The birthrate determines the number of children the agent gets. If, for example, the birthrate is 1.5, the agent produces a single offspring and with a 50 percent chance a second offspring. See equation 4.4 for the exact calculation.

$$numSites = (2CarryingScale + 1)^2 \tag{4.4a}$$

$$K = numSites \times idealNumAgentsSpot \tag{4.4b}$$

$$r = 0.5 \tag{4.4c}$$

$$P = numNeighbors \tag{4.4d}$$

$$PopulationGrowth = \frac{dP}{dt} = rP(1 - \frac{P}{K}) \tag{4.4e}$$

Development of Language

During childhood the child slowly learns language. How this process develops can vary, due to different parameter settings. First of all, the ADULTAGE has consequences on the learning process of a child, because the LEARNSTEPS are spread equally over the child years. The default number of LEARNSTEPS and the ADULTAGE are respectively determined at 100 and 5. This entails every simulation step or year, consists of 20 learn steps or encounters with adults. In a simulation step, the parent may move to another spot and/or other agents move to other spots. The environment of the child has a significant chance of changing. When the ADULTAGE is high, this chance is higher, causing a higher chance on learning from other agents. One can choose to have a child become an adult after a single step. Then the child only, assuming the child learns from others at some time, learns from a single other adult. When the child has an adult age of 18, and the child is able to learn from others at 10 it has 9 years in which there is a significant chance different adult agents are encountered. In this case the child agent is exposed to several potentially different languages. The variety is greater. But if the child agent learns from others, the languages may tend to level out or *converge* more.

In the SPM the child only learns a single language, but one can imagine it is logical to incorporate the learning of multiple languages. But, there are several problems with learning multiple languages. The main problem is that the complexity of the model increases significantly.

Because of these significant problems the model is simplified to a model where agents posses only single languages. At first the child learns only from its parent, and when the child is allowed to learn from others, it checks if the utterance, presented by a non-parent, is too different from its own language. In this project "too different" entails unintelligibility. If this is the case, the child dismisses this adult as a teacher and the next adult agent is checked. This is done until an adult agent is found which offers an utterance not to different from the child´s language. If no other agent is found in the child's neighborhood (which is determined by the parameter LANGUAGEVISION) the parent is presented as a teacher. That child agents only learn from agents with similar languages, reflects the idea proposed by Axelrod [1997], see 2.8.

When is a language too different

Naturally, it is difficult to determine when two languages are too different, but in this thesis a huge simplification is done. The idea has originated simply by viewing several example languages (several constraint orders) with different numbers of active constraints and subjectively judging when these two languages are too different. To determine whether the language of the child differs too much from the utterance presented by the adult, the first thing to do is determining the order of the constraints of the utterance in the language of the child. The second thing to do is calculating the actual difference between the ranking of the constraints of the adult and child agent. Below the pseudo code of the function *calculateDifference.*

Figure 4.5: Pseudocode of the function to calculate the difference between an utterance and the child its language

```
function CALCULATEDIFFERENCE(childOrder, utterance, childActiveConstraints)
    inputs: childOrder
            utterance
            childActiveConstraints
    difference ⟵ 0
    for every constraint C_i in utterance
        if childActiveConstraints not contains C_i
            difference += (utterance.indexOf(C_i) + 1) × 2
        else if (childOrder.indexOf(C_i) is not utterance.indexOf(C_i))
            difference += difference between utterance.indexOf(C_i) and childOrder.indexOf(C_i)
    end for
    return difference
```

The maximum difference possible is calculated using the *numUtteranceConstraints*, with the following equation:

$$n = numUtteranceConstraints \tag{4.5}$$

$$\sum_{i=1}^{n} 2i = n(n+1) \tag{4.6}$$

This maximum allowable difference between two languages is determined at one third of the maximum difference. One third

Nettle simulation

When the NETTLE checkbox is checked, a special parameter setting is loaded. The model which is explained in section 2.7.3, will then implemented. The grid is 7 by 7 with on 20 agents on every site. The agents ages are between 1 and 5 and the adult age is 2, so that only agents of 1 learn language. They learn the average language of the adult agents around them. There is no movement. The difference between the implementation of the model of Nettle in the SPM and the actual model of Nettle, is that the representation of language is different. In the model of Nettle a vowel system is used.

Table 4.2: Parameters

Parameter	Description	Default
GridXSize	The width of the agent grid	5
GridYSize	The height of the agent grid	5
CarryingScale	The scale which determines the size of the neighborhood for all reproductive agents, where the population growth is controlled, see 4.4	0
NumAgents	The initial number of agents	20
MinDeathAge	The minimum age at which an agent can die	17
MaxDeathAge	The maximum age at which an agent can die	17
MinReproduceAge	The minimum age at which an agent can reproduce	6
MaxReproduceAge	The maximum age at which an agent can reproduce	9
AdultAge	The age that a child becomes an adult	5
Mobile	This is true or false; agent are able to move or not	true
MoveChance	Chance for the adult agents to move at every step	0.04
Replacement	true or false: If true, the agents are replaced by a new child	false at death
UniformSpaceDistribution	true or false: the agents are uniformly or randomly distributed over the scape	false
UniformAgentsPerCell	The number of agents at every cell or site	n.v.p.
Nettle	A special boolean. If this is chosen the model will simulate the model of Nettle [1999], see 4.2.2	n.a.
NumConstraints	The number of constraints in the language	10
NumActiveConstraints	The number of active constraints	3
NumUtterenceConstraints	The number of constraints whereof an utterance consists	3
Plasticity	The learn rate of a child	1.0
LanguageScale	The size of the continuous scale where constraints can lie on	35
LearnSteps	The number of learn steps the child gets to learn the language	100
LearnFromOthers	The percentage of its life at which the child can learn from other agents	0.5
LanguageVision	The vision a child has around itself to search for other agents to learn language from	0
Disaster	true or false: If true there will be a disaster, from the decided location agents will die around this location with a certain scale	false
LocationXDisaster	The x location of the heart of the disaster	n.a.
LocationYDisaster	The y location of the heart of the disaster	n.a.
ScaleDisaster	The size or radius of the disaster	n.a.
TimeDisaster	The time at which the disaster will occur	n.a.

Chapter 5

Description of the Measures on the Spatial Population Model

5.1 Measures for Linguistic Diversity

To measure the linguistic diversity in a population and especially while running the simulation of this model several things need to be considered. First, linguistic diversity can simply be measured by counting the number of languages in the population. More precisely, this measure is a measure of language diversity as described in section 2.3. Second, it is important not only to know the number of languages, but also the distribution of the languages over the agents. In other words, what are the proportions of agents per language. If there is one language possessed by many agents, and 5 other languages each possessed by only single agents, this entails lower diversity then when the agents would be more evenly distributed over the languages. Third, it is also important to consider the distance or difference between languages in the population. If the languages are very different the linguistic diversity should be higher than when the languages are close together. This aspect of linguistic diversity is actually referred to by Nettle [1998] as the *phylogenetic diversity* and *structural diversity*, see section 2.3.

In order to measure these three things two measures have been operationalized in the model to calculate the complexity of the linguistic diversity over a group of grammars or languages. First, there is the *simple social entropy* from Shannon [1949], which can cope with the first two. Second, there is the *Hierarchic Social Entropy* developed by Balch [2000], which takes all three things into consideration and therefore is a more complete measure of linguistic diversity. But before covering these measures it is important to show how the distance between two languages is measured.

5.1.1 Measuring the Difference between two Languages

A quantitative measure of the difference between languages is needed for all the other more general measurements and for parts of the visualizations. Therefore it is important to start with this. In this project two types of difference measures are used. First, there is the continuous, second the discrete measure.

Continuous Difference

The *continuous difference* is essentially the average distance between two languages. This average distance is simply calculated by the following equation, 5.1.

$$continuousDifference = \frac{1}{\mathcal{N}_{\mathcal{P}}} \sum_{n=0}^{\mathcal{N}_{\mathcal{C}}} |c_{1,n,t} - c_{2,n,t}| \tag{5.1}$$

This equation simply sums the absolute differences between the rankings of the constraints of the two languages and divides this by the number of active or participating constraints, $\mathcal{N}_{\mathcal{P}}$. $\mathcal{N}_{\mathcal{P}}$ is used because there are $1 - \mathcal{N}_{\mathcal{P}}$ passive constraints that are in both languages 0. Consequently, the difference is only caused by difference between active constraints. The AVERAGEDISTANCE is the average distance between a single constraint in the one language and the same constraint in the other language.

Table 5.1: Ranking of two adult grammars of the same language with the difference between them in the last column

constraints	Ranking Adult 1	Ranking Adult 2	Absolute difference
A	2	10	8
B	4	25	21
C	8	26	18
Total			47

In table 5.1 an example is given of two adult grammars of the same language with its rankings and difference in rankings. The total difference is therefore 47, and the average distance $47/3 = 15\frac{2}{3}$.

The problem with this measure is that it does not differentiate between different constraint orders. In 5.1 the adult grammars are very different on a continuous scale, but on a discrete scale they are equal: [A, B, C].

Discrete Difference

The *discrete difference* can cope with different constraint orders, and has not been used by anyone else. The measure is almost similar to the *Damerau-Levenshtein distance*, [Damerau, 1964]. Strictly speaking, the Damerau-Levenshtein distance is equal to the minimal number of insertions, deletions, substitutions and transpositions needed to transform one string into the other.

The pseudocode of the *discrete difference* between two adult grammars is given in figure 5.1. It is similar to the calculation of difference between an adult and a child grammar, given in 4.5, but there is the subtle difference that this comparison needs to be symmetric. If *grammar1* and *grammar2* are swapped, in every case, the difference needs to be the same. This is not the case in 4.5, where solely the child grammar is compared with the adult grammar. The function is the same, but only the last if statement is new. This statement checks whether the *constraint2* of *grammar2* is present in *grammar1*. Also when a constraint is not present in the other grammar, the difference is $i + 1$ in stead of $(i + 1) * 2$. The maximum possible difference between grammars is the same as in 4.5 and is given in 4.5.

To show how this difference is calculated there follows an example run through the function. Grammar1 is [A, B, C] and grammar2 is [B, D, C]. In the first loop, j is 0, constraint1 is A and constraint2 is B. Grammar2 does not have A, so 0 + 1 is added to the difference. No other condition is satisfied. In the second loop, j is 1, constraint1 is B and constraint2 is D. Grammar2

Figure 5.1: Pseudocode of the function to calculate the discrete difference between two adult grammars

```
function CALCDISCRETEDIFFERENCE(grammar1, grammar2)
    difference ⟵ 0
    for (int i = 0; i ¡ grammar1.size(); i++)
        constraint1 ⟵ grammar1[i]
        constraint2 ⟵ grammar2[i]
        if grammar2 not contains constraint1
            difference += i + 1
        else if (grammar2.indexOf(constraint1) is not i)
            difference += |grammar2.indexOf(constraint1) - i|
        if grammar1 not contains constraint2
            difference += i + 1
    end
    return difference
```

does have B. Now the function checks whether B is **not** in the same position in grammar1 as in grammar 2. This is the case so the displacement, which is one, is added to the difference, making the temporary difference 2. The last **if** statement is satisfied, because grammar1 does not have D. Now 1 + 1 = 2 is added to the difference making it 4. The C occupies the same position in both grammars, so no difference here. The total difference is 4. When the grammars are swapped, the same result results.

Discrete vs. Continuous

Why are there two measures measuring the diversity in two different ways? First, this is because two different measures cause a more complete view of the language diversity. The advantage of the continuous difference measure is that languages are almost always different. The sensitivity is higher. If the order of the constraints in two languages is the same, making the discrete difference zero, but the constraints are a little different in its ranking, the continuous measure observes a slight difference. In such cases it can be concluded that the constraints are close together, otherwise this small difference is not possible, and that the chances of constraint swapping are high or that the chances of conservation are low.

A drawback of the continuous difference measure is that the order of constraints is neglected. It is easily possible that the continuous difference between languages is small, but the order completely different. The order of constraints is important, because the theory of language in this project is that the order of constraints determines the language and not the ranking. To determine if two languages are different, theoretically, the order of constraints is decisive because these determine the eventual optimal forms, see section 3.1.

An advantage of the discrete difference measure is that far less languages are different. This makes the visualization, explained in .1, more synoptic, because there is less coloring and language regions are more precisely determined.

It can be concluded that the discrete difference is considered more important in determining the language diversity first, because it displays the theoretically based discrete difference, and second, because it is more synoptic. But, the continuous difference is certainly of importance too.

5.1.2 Simple Social Entropy

Shannon [1949] introduced the measure of *information entropy* with the goal to quantify information source uncertainty. Interesting is that this measure is also useful in the measurement of societal diversity, and in this case linguistic diversity. A passage from E.O. Wilson's book *The Diversity of Life* illustrates this, cited from Balch [2000, page 211].

> Suppose that we encounter a fauna of butterflies consisting of 1 million individuals divided into 100 species. Say one of the species is extremely abundant, represented by 990,000 individuals, and each of the other species therefore comprises an average of about 100 individuals. One hundred species are present but, as we walk along the forest paths and across the fields, we encounter the abundant butterfly most of the time and each of the other species only rarely ... In a nearby locality we encounter a second butterfly fauna, comprising the same 100 species, but this time all are equally abundant, represented by 10,000 individuals each. This is a fauna of high equatability, in fact the highest possible. Intuitively we feel that the high-equatability fauna is the more diverse of the two, since each butterfly encountered in turn is less predictable and therefore gives us more information on average.

This illustrated the idea that societies with members equally distributed over species are the most diverse. This social entropy measure is adapted to a calculation of simple social entropy in robot groups in Balch [2000]. An elaborate description towards the social entropy measure is given here. The notation used is copied from Balch [2000, page 4].

- $\mathcal{R}$ is a society of N agents with $\mathcal{R} = \{r_1, r_2, r_3, ..., r_N\}$
- $\mathcal{C}$ is a classification of $\mathcal{R}$ into M possibly overlapping subsets.
- c_i is an individual subset of $\mathcal{C}$ with $\mathcal{R} = \{c_1, c_2, c_3, ..., c_M\}$
- $p_i = \dfrac{|c_i|}{\sum_{j=1}^{M} |c_j|}$ is the proportion of agents in the ith subset.

Next, the information entropy of a system X is:

$$H(X) = -K \sum_{i=1}^{M} p_i log_2(p_i) \tag{5.2}$$

K is a positive constant, where Shannon [1949] puts this at $K = 1$. In this project the notation needs to be adjusted. see below:

- $\mathcal{R}$ is a group of adult agents of N agents with $\mathcal{R} = \{r_1, r_2, r_3, ..., r_N\}$
- $\mathcal{A}$ are the adult languages or clusters of languages of size O with languages $\mathcal{A} = \{a_1, a_2, a_3, ..., a_O\}$
- $\mathcal{C}$ is a classification of $\mathcal{A}$ into M non-overlapping subsets.
- c_i is an individual subset of $\mathcal{C}$ with $\mathcal{C} = \{c_1, c_2, c_3, ..., c_M\}$
- $\dfrac{|c_i|}{\sum_{j=1}^{M} |c_j|}$ is the proportion of adult languages in the ith subset; $\sum p_i = 1$.

So, every adult agent has a language, and consequently falls into a subset c_i. Only the adult languages are evaluated, because the languages of the children are not fully developed yet.

The measured diversity of the language population should reflect the number of languages and the distribution of adult agents over these languages. M is the number of languages and the

p_is are the proportions. An example of a diversity metric can be $H(p_1, p_2, p_3, ..., p_M)$. If there are 4 languages, $M = 4$, in the adult population and the proportions are $\frac{2}{12}, \frac{4}{12}, \frac{5}{12}, \frac{1}{12}$, this can be written as $H(\frac{2}{12}, \frac{4}{12}, \frac{5}{12}, \frac{1}{12})$.

Because this project is about an agent society the information entropy formula is rewritten into the *simple social entropy* of agent society $\mathcal{R}_a$.

$$H(\mathcal{R}_a) = -K \sum_{i=1}^{M} p_i log_2(p_i) \tag{5.3}$$

There are 6 important properties which H has which can be found in Balch [2000]. Example evaluations are given in Balch [2000]. Following is an example is given of a calculation of the entropy of a language population.

5.1.3 Limitations of Simple Social Entropy

Several limitations of *Simple Social Entropy* are mentioned in Balch [2000]. These will be shortly mentioned here.

Occam's Razor

The simple social entropy is a good abstract measure for the degree of linguistic diversity, but because it is a single value a lot of information is lost. It does not tell how many classes of languages there are or how many adult agents in a single class. A single entropy value can represent several totally different compositions of language populations.

Lack of Sensitivity to the Degree of Difference between languages

This is a more serious limitation according to Balch [2000]. The simple social entropy does not give a lot of information about the degree of difference between grammars. A given grammar is the same as some other grammar or not. If grammars are very different, measured by the discrete difference (section 5.1), one could say these grammars display a higher value of linguistic diversity, than when grammars are close together. An example is given where the diversity of a set of agents is shown in a two-dimensional space.

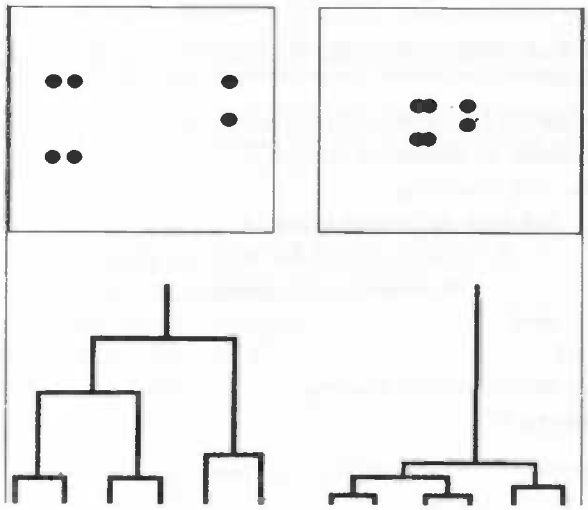
5.1.4 Hierarchic Social Entropy

Because simple social entropy has several limitations, *Hierarchical Social Entropy* is introduced [Balch, 2000]. *Hierarchical Social Entropy* is a more complex measure, which can give a significantly more detailed picture about the linguistic diversity in this model. More precisely, it combines the simple social entropy, with the maximum differences between languages on a continuous scale.

The *Hierarchical Social Entropy* is a value which is constructed by applying a hierarchical clustering algorithm to several taxonomic levels of maximum differences between languages. The goal is to build a *taxonomic tree* that reflects the hierarchic distribution of the elements in the system. Closely related languages should be classified together at the bottom of the tree. Higher in the tree, similarities between groups make these converge into a parent group. This converges until there is only a single branch left, containing all the languages. A taxonomic tree can be graphically displayed in a *dendrogram*.

These dendrograms are very useful in discovering a detailed picture of the linguistic diversity in a society. In 5.2 an example from Balch [2000] is given. These are pictures from two societies, above the 2-dimensional display of differences between elements are shown and below the dendrograms. The relative arrangements of the elements is the same, but the right society is compacter than the left entailing less diversity. This is reflected in the respective dendrograms, where the vertical axis represents the difference between elements.

Figure 5.2: Example dendrograms



The *hierarchical clustering algorithm*, divides a society into subsets of behaviorally equivalent agents at a particular taxonomic level. In the case of this project this means that the language population is divided into linguistically equivalent agents at a particular maximum difference between the languages. This way, with different maximum difference, it is possible to observe the language structure at different levels of difference. These different levels can all be displayed in a dendrogram

The comparison can be made with the linguistic diversity in real life. Dialects would be grouped together in lower taxonomic levels, and large language families in higher ones. Therefore, this is a measure of Phylogenetic Diversity (introduced in 2.3).

Hierarchic Clustering Algorithm

This clustering method is hierarchic because any lower ranked taxon is also a member of a higher ranked taxon. A clustering algorithm clusters the adult languages or clusters of adult languages based on the maximum difference, h, between elements. These can be clusters, because the clustering method can, at a certain moment, be applied at a high level where the direct lower level already consists of clusters of languages.

The measure used here assumes that the clusters are non-overlapping. The main reason for this is the high computation cost to have a well-formed overlapping clustering algorithm.

The language population consists of children and of adults. This measure is applied to the languages of the adults. $\mathcal{A}$ are the adult languages or clusters of languages of O, which will be clustered. The adult languages or clusters will be divided into M non-overlapping clusters $\mathcal{C}$ with $\mathcal{C} = \{c_1, c_2, c_3, ..., c_M\}$.

- $\mathcal{R}$ is a group of adult agents of N agents with $\mathcal{R} = \{r_1, r_2, r_3, ..., r_N\}$
- $\mathcal{A}$ are the adult languages or clusters of languages of size O with languages $\mathcal{A} = \{a_1, a_2, a_3, ..., a_O\}$
- $\mathcal{C}$ is a classification of $\mathcal{A}$ into M non-overlapping subsets.

- c_i is an individual subset of $\mathcal{C}$ with $\mathcal{C} = \{c_1, c_2, c_3, ..., c_M\}$
- $\frac{|c_i|}{\sum_{j=1}^{M} |c_j|}$ is the proportion of adult languages in the ith subset; $\sum p_i = 1$.

The algorithm, copied from Balch [2000], at level h works as follows: Initially, a O number of

Figure 5.3: Pseudocode of the function to create clusters at a certain maximum difference, h

```
function ClusteringAlgorithm(A, h)
    initialize O clusters with c_i = a_i
    for each cluster c_i:
        for each a_j (except a_i) in A:
            if (D(a_j, a_k) ≤ h) for every a_k already in c_i
               add element a_j to cluster c_i
        end
    end
    discard redundant clusters.
    return C
```

clusters is initialized with the to be clustered elements $\mathcal{A}$. This means c_1 is initialized with a_1 and c_M is initialized with a_O. Then for each element c_i, iterate over all the elements in $\mathcal{A}$. If a single a_i is not more different than h, with every other element the current cluster c_i already contains, this a_i will be added to the current cluster. At the end of the for loop there are N clusters, each with one or more elements of $(\mathcal{A})$. But it is possible, that one or more groups of clusters contain the same elements. The last statement takes care of this by removing redundant clusters.

The elements a_i can be clusters of languages. In this case the average of the differences between the languages in element a_1 and in element a_2 is used as the difference.

In this project a hierarchical non-overlapping clustering method is used.

Building of the taxonomic tree

The clustering algorithm is used to compose the clusters at a certain taxonomic level h. The goal is to build a complete taxonomic tree.

- $\mathcal{T}$ is a taxonomic tree of variable size N of taxonomic levels $\mathcal{T} = \{t_1, t_2, t_3, ..., t_N\}$
- Every t_i consists of a $\mathcal{C}$, which is a classification of $\mathcal{A}$ into M non-overlapping subsets and has a certain h and a certain social entropy S. $t_i = \{\mathcal{C}, h, S\}$

When the CreateTaxonomicTree function is called, a taxonomic tree is created. If there is only a single language in the current society, the creation of the taxonomic tree simply consists of a single taxonomic level which consists of all the languages. Else, the BinarySearch function is called to construct the hierarchic tree. The BinarySearch function recursively searches where the taxonomic levels are located on a scale of h between 0 and 1. The function locates the levels accurately.

BinarySearch begins by composing the set of clusters at h value 0.5, with the ClusteringAlgorithm. First, this set can consist of fewer clusters than the previous taxonomic level, which means a new taxonomic level is found. Nonetheless, it could mean that there are more taxonomic levels at lower h values. Moreover, it is important to accurately locate where the transition to this new taxonomic level lies. The only thing known is that at an h value of 0.5 there are fewer clusters than at an h value of 0. But what happens in between is unknown. Second, it could also be that at h 0.5 no new taxonomic level is found. If there are equal clusters, the h is increased, else decreased by half the previous increment/decrement. In this case this is 0.25, because the previous increment was from 0 to 0.5. This process is continued until a h value is reached where there are fewer clusters and the previous increment/decrement is lower than a certain *small value*. The last is to have high accuracy. Then a new taxonomic level is added to the taxonomic tree. Then the function is called again but now the starting h is the previous h. This process continues until the last taxonomic level, consisting of one cluster, is found.

Figure 5.4: Pseudocode of the function to create the taxonomic tree

```
function CREATETAXONOMICTREE(A)
    numLanguages ⟵ size(A)
    if numLanguages == 1
        t1 ⟵ {A, 0, 0}
        T ⟵ {t1}
    else
        lowH ⟵ 0
        h ⟵ 0.5
        prevH ⟵ 0
        numPrevClust ⟵ numLanguages
        T ⟵ ∅
        BINARYSEARCH(A, T, lowH, h, prevH, numLanguages)
```

Calculation of Hierarchic Social Entropy

With the taxonomic tree $\mathcal{T}$ the *hierarchic social entropy* can be calculated. In the previous two paragraphs it is explained that the taxonomic level depends on the clustering at a certain h. The following notation is used for this, [Balch, 2000]:

$$H(\mathcal{R}, h) = \text{entropy for the clustering of } \mathcal{R} \text{ at taxonomic level } h \tag{5.4}$$

Now there is a need for calculating the total entropy or *hierarchic social entropy* of the clustering at all taxonomic levels. Balch [2000] solves this by taking the area under the entropy plot. The *hierarchic social entropy* is defined as follows:

$$S(\mathcal{R}) = \int_0^\infty H(\mathcal{R}, h)dh \tag{5.5}$$

In Balch [2000] the complete calculation is given, but this function simply calculates the area under the entropy plot. For example, see figure 5.6, where the H(1/3, 1/3, 1/3), H(2/3, 1/3), H(3/3), are the taxonomic levels and the proportions of the clusters at the specific taxonomic level. The first taxonomic level is for $0 \le h < x$ and has a simple social entropy of 1.585 the second 0.811 and the last 0. The *hierarchic social entropy* is for a x of 0.1, 0.2396.

The Problems with the Hierarchical Social Entropy

In this project the HSE has been used as the main measure of linguistic diversity. However, there are some problems with it that have to be noted and cause a slight inaccuracy in this measure. This problem is due to the *non-overlapping* feature. Because the clusters can not be overlapping, the hierarchic tree is almost never completely accurate in distinguishing the right differences between languages. An example of this is shown by the following. The dendrogram in 5.7 is conceived from an ad hoc run. The top five languages of the hierarchic tree are:

First branch •
- $[A1, A8, A3]$
- $[A1, A3, A8]$

Second branch •
- $[A7, A8, A3]$
- $[A5, A8, A3]$
- $[A9, A8, A3]$

The two branches are later joined into a single branch, indicating that there should be more differences between the two branches than within the branches. But the first language is just as different from the second language in the first branch as the three languages in the second branch. Because the clusters are not allowed to overlap, the second branch is not accompanied by the first language in the first branch, while it should. Although this is a problem, it is a slight one, because the differences between languages in a society are still deducted precisely enough to indicate the amount of linguistic diversity.

5.1.5 Spatial Distribution of the Languages

In the previous measures the *spatial distribution* of the languages has been neglected. Nonetheless it is fairly important to know whether language groups are close together or spread over the scape. Therefore a new complex measure for the spatial distribution is made up with help of the taxonomic tree. The taxonomic tree provides hierarchic information about the languages, about which groups form dialects and which groups form families. When there are quantative values for the spatial distribution of these different groups, these can be summed into a single value, which provides a sensitive measure for the distribution.

This single value for the spatial distribution of the languages is calculated by summing the spatial distributions of all the clusters in the taxonomic tree and dividing it by the number of clusters. In figure 5.2 there are for example 11 clusters: 6 leafs, then 3 pairs of leaves merge into 3 branches, then the right two pairs join in a single branch, and lastly the remaining branches merge into a single cluster. $6+3+1+1$ result in 11 clusters. The spatial distribution of a cluster is calculated by taking the *center of mass* L of the locations of all M languages in this cluster and then calculating the standard deviation of all the locations from the center of mass. The mass of all languages is the same, so only the locations l_i are needed:

$$L_c = \frac{1}{M} \sum l_i \tag{5.6}$$

Figure 5.5: Pseudocode of the function to binary search

```
function BinarySearch(*𝒜, 𝒯, lowH, h, prevH, numPrevClust)
  inputs: lowH, the value which is in the current search the lowest value
          h, the current h
          prevH, the h value of the previous taxonomic level
          numPrevClust, the number of clusters in the previous taxonomic level
  if Size(𝒯) == 0
    𝒞 ←— ClusteringAlgorithm(𝒜, h)
  endif
  else
    𝒞 ←— ClusteringAlgorithm(𝒯.t_n.𝒞, h)
  if abs(prevH − h) < smallvalue and Size(𝒞) == 1
    if (size)(𝒯) == 0
      add new t{𝒜, h, entropy} to 𝒯
    endif
    add new t {𝒞, 1, entropy*} to 𝒯
    return
  endif
  else if 𝒞_M == numPrevClust
    BinarySearch(𝒯, lowH, h + abs(prevH − h)/2, h, numPrevClust)
    return
  endif
  else if abs(prevH − h) > smallvalue
    BinarySearch(𝒯, lowH, h − abs(prevH − h)/2, h, numPrevClust)
    return
  endif
  else
    add t{t_n.𝒞, prevH, entropy*} to 𝒯
    BinarySearch(𝒯, h, h + (1 − h)/2, h, 𝒞_M)
    return
  endif
```

* The entropy is calculated by equation 5.3

Figure 5.6: The Simple Entropy graph of the several taxonomic levels in the hierarchic tree

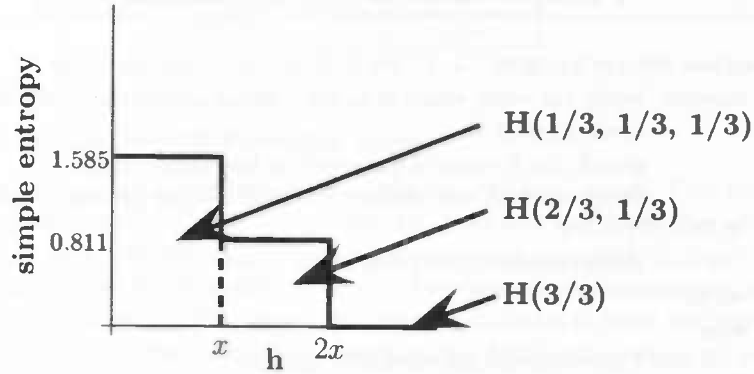
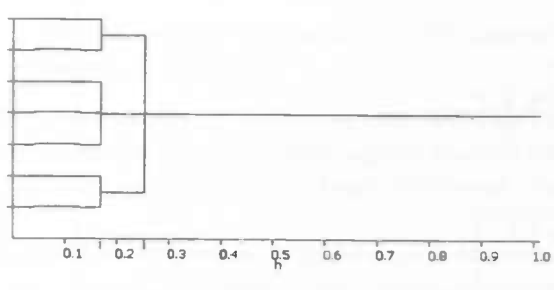


Figure 5.7: An example of a dendrogram to show the problems with it



Chapter 6

Experiments and Results with the Individual Language Model

A number of experiments have been done on the Individual Language Model (ILM) to get a feeling of what kind of influence the several parameters have on the development of language. First, there are the experiments on a single learning cycle between a single adult and a single child. Then, tests have been done on the *Iterated Learning Model* (IL).

6.1 Tests on a Single Learning Cycle between Adult and Child

The general goal is to investigate influences on linguistic diversity. In the following tests influences on the degree of language change are tested. The language change is the difference between the language the child has learned from a single adult and the language of the adult. Several parameter values are tested to get a good idea about which combinations of parameters are interesting.

The test consists of a number of learning moments between the adult and child agent. The first set of tests consist of recordings of the continuous difference between the languages with several values of several parameters over time, see section 5.1.1 for an explanation of this measure. The second set of tests consist of recordings of the percentage of innovations and the discrete difference between languages with several values of several parameters.

6.1.1 Continuous Difference Tests

The influence of the following parameters on the development of the continuous difference are tested:

- PLASTICITY values: The learning rate by which the child learns the language.
- *Percentage* of utterance constraints: The percentage of active constraints used for creating an utterance.
- The LANGUAGESCALE: The size of the continuous scale on which the constraints lie.

In all tests the parameters have the following standard values:

Table 6.1: ParametersDefault

Parameter	Default
NumConstraints	10
NumActiveConstraints	3
NumUtterenceConstraints	3
Plasticity	0.1
LanguageScale	50
LearnSteps	1000

Continuous Difference with Several Plasticity Values

The first test is about the continuous difference with different plasticity values. Five tests were done with the plasticity values (p), 0.05, 0.1, 0.3, 0.5 and 1.0. For every test there where 100 runs. And for every run there where 1000 learning moments between adult and child. In figure 6.1, the 1st Quartile, the median and the 3rd Quartile are shown. Two general observations can be done. First, with *higher* plasticity values the child needs less learning moments to learn the language as well as it can. With plasticity 0.1, the child needs approximately 500 learning moments and with plasticity 1.0 around 50. Second, the learning accuracy of the language increases with a lower plasticity. The 1st quartile and the median of the graph with plasticity 0.1 lie lower than these of the graph with plasticity 1.0. The 3rd quartiles are almost equal but high and around 8, indicating that for every plasticity value the differences can be equally high. The graph with plasticity 1.0 shows that the median difference nears 2. This is the median difference a single constraint in the child language has with the same constraint in the adult language. 2 is the standard deviation of the normal distribution of a constraint and 1/25 of the maximum distance of 50.

Continuous Difference with Several Utterance Constraint Percentages

The second test is about the influence of the percentage of active constraints in the utterance on the development of the continuous difference between the adult and child language. Here only three tests have been done: the utterance consisting out of a single, two and three constraints, randomly chosen out of three active constraints. The results are displayed in figure 6.2. The main observation is this:

> A larger number of constraints in the utterance result in an increase of the speed of learning the language

This is because more is learned at every learning moment. If only one constraint is presented at every utterance, the child simply adjusts this one constraint. If three constraints are presented at an utterance, the child adjusts three constraints. There are no significant accuracy differences.

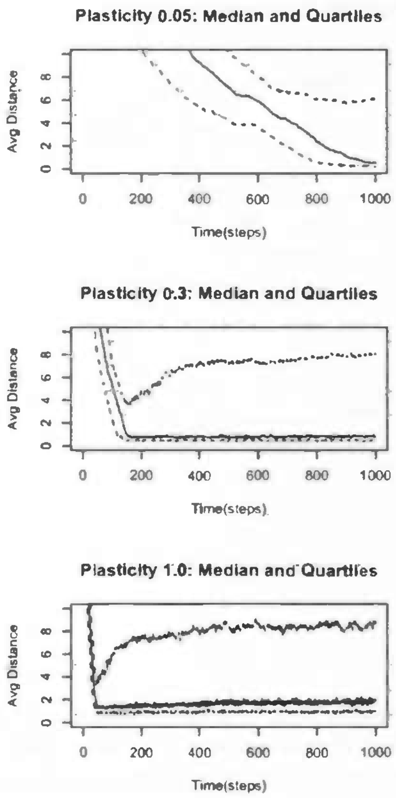
Continuous Difference with Several Language Scale Sizes

The third and last test with continuous difference is about the influence of the size of the language scale on the continuous difference. Five tests with the following scale sizes have been performed: 10, 20, 35, 50 and 75.

Now three general observations can be extracted from figure 6.3:

1. A smaller scale results in faster learning
2. A larger scale results in higher accuracy

Figure 6.1: 5 graphs displaying the three quartiles of the continuous difference the child language and the adult language have with different plasticity values



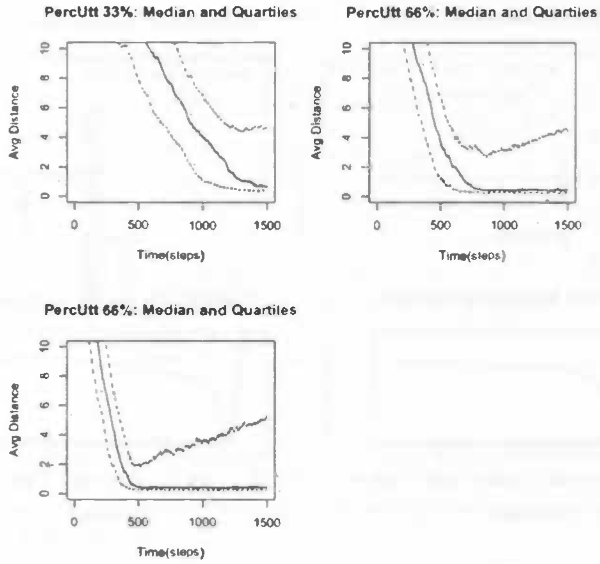
3. Especially with scale 10 and less so with scale 20 after reaching the valley the child unlearns the language

6.1.2 Innovation and Discrete Difference Tests

The next set of tests is about the percentage of *innovation* and the *discrete difference* between the language of the adult and the learned language of the child, under different conditions. An innovation occurs when the child has learned a language containing one or more new constraint(s), which was/were not present in the language of its teacher. The percentage in a single run is calculated by comparing the child language with the adult language: the percentage of new constraints in the child language. E.g. the adult language is [A1, A2, A3, A4] and the child language is [A1, A2, A5, A6], this means the percentage of innovation is 50%.

The standard values of the parameters stay the same, similar as with the continuous difference tests. The parameters tested are scale, plasticity and the number of learning moments. For every test, 3000 runs have been performed. In the plots in figure 6.4 a boxplot and a histogram with the averages and the standard errors are shown. For the measurement of the percentage of innovation this means that if there are for example 60 runs where the child has a language with 50% innovation, the percentage is 1%.

Figure 6.2: The continuous difference between the child language and the adult language with different utterance percentages



Scale

The general trend in the scale tests is that an increasing scale results in first a fast decrease and then every time a slower decrease in the percentage of innovations. The same holds for the average discrete difference. From these two observations, it can be concluded that the *discontinuity of language transmission* increases with a smaller scale.

Plasticity

The general trend in the plasticity tests is that an increasing plasticity results in an increasing percentage of innovations and discrete difference.

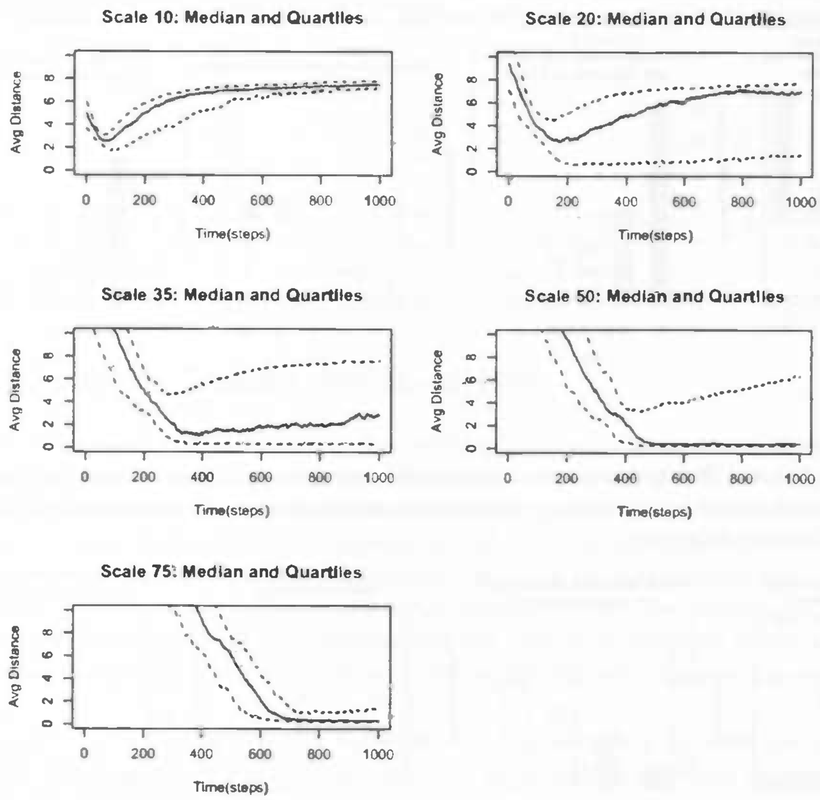
Learning Moments

The results of this test requires a somewhat more elaborate discussion. The percentage of innovations increases linearly with increasing iterations. This makes sense because over time the chances grow for innovations to occur. But the discrete difference showed the opposite effect: increasing number of iterations caused an exponential decrease in the discrete difference, stabilizing from 500 iterations.

6.2 Tests on the Iterated Learning Model

The next tests are on the *Iterated Learning Model* where several values of several parameters are tested over generations. The iterated learning model is discussed in 4.1.6. The IL is simply an iteration over generations of agents. It starts with a single adult and child agent. The child becomes adult and teaches the language to a new child. This continues until a stop is forced. Two tests have been performed with the ILM:

Figure 6.3: The continuous difference between the child language and the adult language with different language scale sizes



1. The Continuous Difference. Section 5.1.1
2. The Discrete Difference. Section 5.1.1

The tests track the continuous and discrete difference between the language of the adult at the current generation with the starting language. In the figures 6.7 a growing line can be seen for the average distance, the percentage of innovations and the discrete difference which entails the following statement which confirms the hypothesis in 4.1.6:

> Over the generations the languages will differentiate more and more from the initial language of the first adult.

The continuous and discrete difference both increase over the generations.

6.3 Discussion

The results where the continuous measure is used show that language is learned to a certain continuous difference, depending on the parameter values. A low plasticity (low degree of imperfect learning) and a large scale (low degree of imperfect performance) result in a low continuous difference. This is only emphasized by the innovation and discrete difference results, where an

Figure 6.4: Boxplots and Histograms of perc. innovations and discrete difference with scales. The boxplots do not show much. The first boxplot shows a hint of a decrease in the beginning but later on they are equal. The average and error bars show a more clearer decrease in the percentage of innovations just as the second

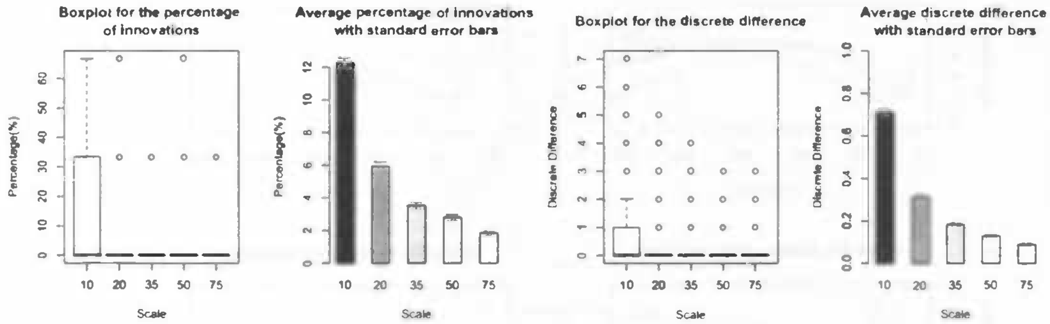
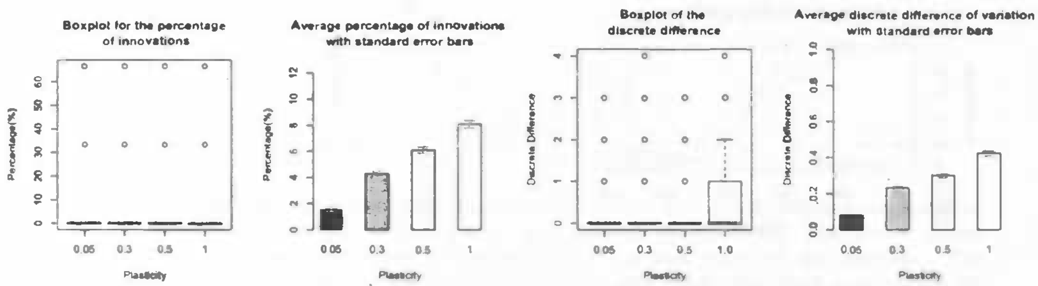


Figure 6.5: Boxplots and Histograms of perc. innovations and discrete difference with plasticity. The same can be said here as in the previous figure, except that there is an increase of innovations and discrete difference



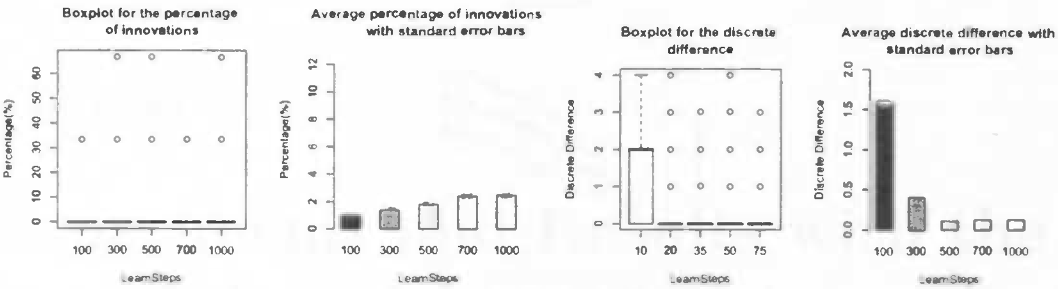
increasing scale and a decreasing plasticity resulted in a decrease of the difference. The increasing LANGUAGESCALE is directly explainable for causing a decreasing difference because it is synonymous to a decrease in noise. The speed of learning, or the number of learning moments necessary to learn the language best is fastest when the plasticity is high, the utterance constraint percentage is high and the scale is low. The tests with the IL show that a single line of generation following generation results in an ever increasing difference.

The results show that this language model, based on SOT and MGLA, is a good learning mechanism of language and is capable of producing variations and innovations in the learned language. This partly answers the second part of the research question:

Research question part II Will the application of the modern theory of language, Stochastic Optimality Theory, work properly and result into new conclusions?

Moreover, the speed and the accuracy by which languages are learned can be easily influenced by changing the certain parameter values.

Figure 6.6: Boxplots and Histograms of perc. innovations and discrete difference with varying numbers of learning moments.



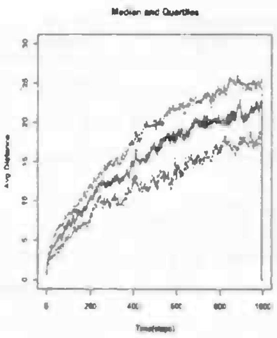
6.4 Parameter Values for the SPM

These tests on the ILM provide insight in the nature of language change in the model and the reaction on several parameters. Moreover, default parameter values for the SPM can be determined.

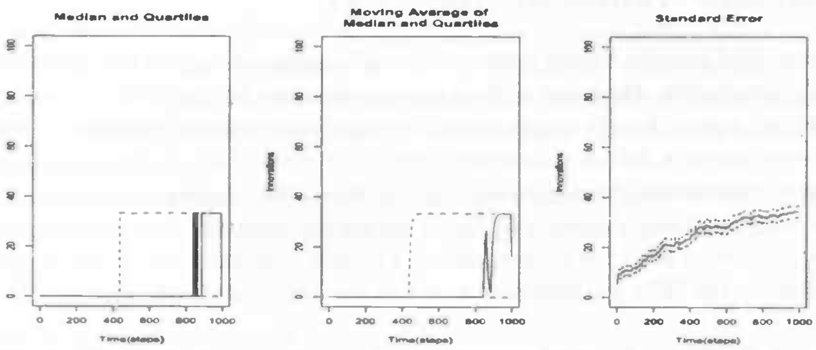
The plasticity results for the continuous difference, percentage of innovations and the discrete difference conclude to a default parameter value of 1.0. First of all, faster learning is preferred because this saves computation time. The accuracy by which the language is learned according to the continuous difference tests is lower with higher plasticity. Also, the percentage of innovations and the discrete difference is higher with plasticity 1.0. But it is uncertain if this is unwished or wished for, because in the SPM the influence of this on the linguistic diversity has yet to be tested upon.

Because the number of learning moments that are necessary to learn the language well enough with default plasticity value 1.0 is 50, but to be safe the default number of learning moments for the SPM is 100.

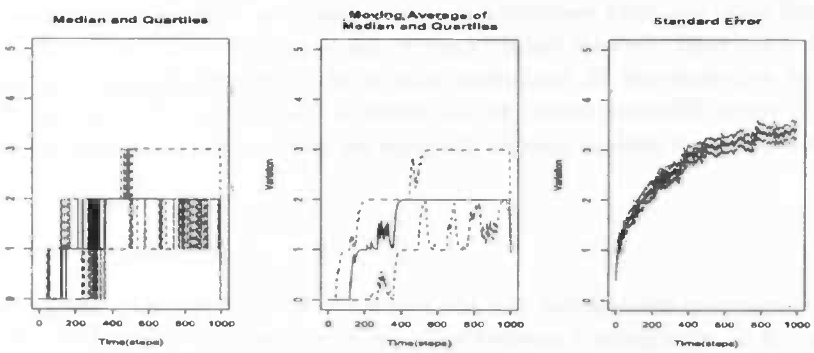
The scale results conclude to a default value of 35. From the continuous difference results it can be concluded that the scale needs to be reasonably large to have a high accuracy learning. With a too small scale the child eventually is able to learn the language worse than with the initial language of the child. Because the accuracy is almost equal for scale size 35, 50 and 75, the learning is fastest with scale size 35, the default scale value is determined at 35 for the SPM. The innovation and discrete difference results do not influence this decision. To conclude, the degree of imperfect performance is average because the scale lies between the lowest and highest tested scale.



(a) Average Distance



(b) Percentage of innovations over the generations: The first graph displays the standard median and quartiles but the lines are abruptly moving up and down. Therefore a second graph has been produced where a moving average has been taken of the quartiles. This produces more curves in the lines, which enables a clearer graph. The third graph shows the average and the standard error, which shows a growing curve



(c) Discrete Difference over the generations: for the same reason in the previous graph, the moving average has been taken of the quartiles. Here, it is more useful, because the contrast between the two graphs is even larger. In the second graph all quartiles grow to values of approximately 2, 2 and 3

Figure 6.7: Results of the Iterated Learning Model

Chapter 7

Experiments and Results with the Spatial Population Model

After the successful (because the representation and transmission work fine and several default parameter values for the SPM have been determined) experiments with the Individual Language Model, experiments have been performed on the Spatial Language Model, described in section 4.2. The default values for the parameters PLASTICITY, LANGUAGESCALE, NUMUTTERANCE-CONSTRAINTS and LEARNSTEPS have been determined and with this the default values are shown in table 4.2.

The goal of these experiments is to provide answers to the research questions. In answering the first research question, the first thing to determine is when one can speak of several reasonable large groups with more languages. The minimum size of a single group is determined at 20, which is the average number of a hunter-gatherer band. When there are 200 agents in the population there can be a maximum of 10 languages. When there are more languages there is too much linguistic diversity. The second thing to determine is: when the linguistic diversity has stabilized. Several measures of the model need to be stable in order to claim stable diversity. These measures are discussed in chapter 5 and chapter .1. These are the number of languages, the number of constraints, the proportions of the languages and the taxonomic tree or dendrogram, which together determine the linguistic diversity neglecting the spatial distribution of the languages. The last two measures are reflected in the hierarchic social entropy measure. The spatial distribution of the languages is also important. When a language group is spread all around the space the spatial distribution is high and when a language group is located at one site it is low. Several experiments are accompanied by these measures.

First there is a general description of the development of the model over the generations, whereafter the experiments and results are presented.

7.1 Development of the Model over the Generations

As the simulation progresses, agents die, generations follow upon generation and languages develop over the generations.

The model generally starts with a small number of agents, whereafter as time passes by, more agents occupy the space. The population growth is controlled by the population growth control mechanism. The diffusion mechanism makes sure the agents spread relatively equal over space. More important is that the languages change. At the start every adult has a single language. As soon as children are born, they learn language from their parent and others. When the child agent reaches adulthood it is almost certain its language deviates slightly from the previous language. This presently adult agent probably reproduces, and its children learn the language slightly differently. This process continues generation after generation.

In 4.1.6 the IL is treated where generations follow up generations. The difference is that in the IL, single agents follow after single agents. In the SPM there are multiple agents acting at the same time. In the IL it seems obvious that language gets in a certain drift of change, where the language a number of generations after generally differ highly from the starting language. The languages in a reasonably far future in the SPM are generally assumed to be largely different from the starting languages.

In the SPM the adult and child agents move around. Children learn from other agents and consequently mix languages. The degree of mixing is influenced by the setting of the language parameters, but a quantity of mixing generally results.

7.2 Experiments

Two sets of experiments have been done: an exploration of the simulation and a more sensitive study. The exploration of the simulation consists of testing the effect on linguistic diversity of several parameters, by using the default values and then changing a single parameter. The sensitive study is aimed at precisely determining the minimum requirements to cause and then stabilize linguistic diversity.

Every experiment has some fixed settings. An experiment consists of 50 runs of 500 iterations. For every experiment data has been stored for every step in every run about:

- The number of adults
- The number of languages present in the population
- The number of constraints present in all the languages of the population
- The number of innovations that have occurred in the run up till the present iteration
- The continuous local diversity
- The discrete local diversity
- The continuous between diversity
- The discrete between diversity
- The simple social entropy
- The hierarchic social entropy

From this data, graphs have been made displaying the 1st, 2nd and 3rd quartile of the 50 runs over 500 iterations. The 500 iterations are chosen, first because it seems after ad hoc testing that after 500 iterations the language population has stabilized, both in number of agents as in the linguistic diversity, and second because with the current age of death of all agents there are 20 generations.

The data need some explanation. The *continuous local diversity* is explained best by looking at the visualization of the continuous diversity space explained in .1.1. The space displays the continuous local diversity at every site. The data that has been collected is the average of this of all the sites. The *discrete local diversity* is almost the same, with the only difference that the measure is discrete. The *continuous between diversity* measures the average difference between sites. This can also be nicely illustrated with the visualization in .1.1. In between sites small rectangles are displayed to show the difference between sites. The *discrete between diversity* is again the discrete variant.

7.2.1 The Exploration of the Simulation

Table 7.1: ParameterValues

Parameter	Description	Values
GridXSize GridYSize	The width and height of the agent grid	3, 5, 7, 9
AdultAge	The age that a child becomes an adult	1, 3, 5, 7, 9
Mobile	This is true or false; agent are able to move or not	true
IdealNumAgents	The ideal number of agents on a site, it indirectly influences the birthrate of reproductive agents	6, 10, 14, 18, 22
NumActiveConstraints	The number of active constraints	2, 3, 4
Plasticity	The learn rate of a child	0.1, 0.3, 0.5, 0.7, 0.9
LearnFromOthers	The percentage of its life at which the child can learn from other agents	0, 0.25, 0.5, 0.75, 1.0
LanguageVision	The vision a child has around itself to search for other agents to learn language from	0, 1, 2, 3

The exploration starts with experiments where all the parameters have the default value and every time a single parameter is varied. Table 7.1 shows the values that are tested. After this, two other experiments have been performed. The first is where the LANGUAGEVISION is varied from 0 to 2 and MOBILE is turned of. This is to test if a higher LANGUAGEVISION decreases the linguistic diversity. The second is where the LANGUAGEVISION is 1 and the LEARNFROMOTHERS is tested on values 0.25, 0.5, 0.75 and 1.0. This test is performed to check if a LANGUAGEVISION of 0 results in a lower linguistic diversity when the child learns from others than with a LANGUAGEVISION of 1. When the child learns from adults in a larger neighborhood it seems logical that it learns a more average version of the language.

7.2.2 Results of the Exploration

First of all it is important to know that only important results are shown.

There are 4 parameters that have proved to make a difference when they are varied: LANGUAGEVISION, LEARNFROMOTHERS, PLASTICITY AND NUMACTIVECONSTRAINTS. It is not necessary to show all graphs to show that there are differences. Only the graphs that resulted

In figures 7.1, 7.2 and 7.3 the graphs for the number of languages, the number of constraints in the world and the hierarchic social entropy (HSE) are given for the LANGUAGEVISION values: 0,1,2,3. This information is sufficient to show that a low LANGUAGEVISION results in a higher linguistic diversity. The leftmost graph in figure 7.1 shows that after an initial increase of the number of languages in the population it stabilizes at around step 150. The median is almost firmly positioned at 5 afterwards. The 1st quartile is steadily 1 lower than the median at 4. The 3rd quartile is however placed high oscillating around 8. This means the distribution is not a normal distribution, but the peak of the distribution has a fast increase from 4 to 5 and than a slower decrease to around 8. This form of distribution is also visible in the graphs with LanguageVision 1. While the LanguageVision increases the general observation is that the trends in the graphs stabilize at a lower number of languages and that the 3rd quartiles decrease to a lower value. There is not a significant difference between the trends of LanguageVision 2 and 3, because the quartiles stabilize all around the same value. In figure 7.2 the main difference can be seen between the graphs in the leftmost and the two on the right, respectively between LanguageVision 0 and 2, 3. In the first the trend is, after stabilizing at around step 150, the median at 4 constraints and the 1st and 3st quartile 1 lower and higher. In the two graphs on the right, the median stabilizes at 3 and the 3rd quartile decreases to 3. The most important observations can be done from the HSE graphs. The differences are clearly visible. With a LanguageVision of 0 the HSE median stabilizes at 0.3, the 3rd quartile is around 0.6 and the 1st around 0.2. Again the main difference is visible between the graph of LanguageVision 0 and LanguageVision 2 and 3. In the latter the 1st, 2nd and 3rd quartiles are eventually all stabilizing at 0.17. The differences are then significant because the 1st quartile of the LanguageVision 0 graph is higher than the 3rd quartile of the LanguageVision 2 and 3 graph.

These results suggest that a LanguageVision of 0 causes a higher linguistic diversity over time than with especially LanguageVision 2 and 3. Secondary to this there is no significant difference to detect between LanguageVision 2 and 3.

Figure 7.1: Exploration: LanguageVision (0,1,2,3) NumLanguages. The trends show a decrease in the number of languages with increasing LanguageVision, although there is no difference between the last two

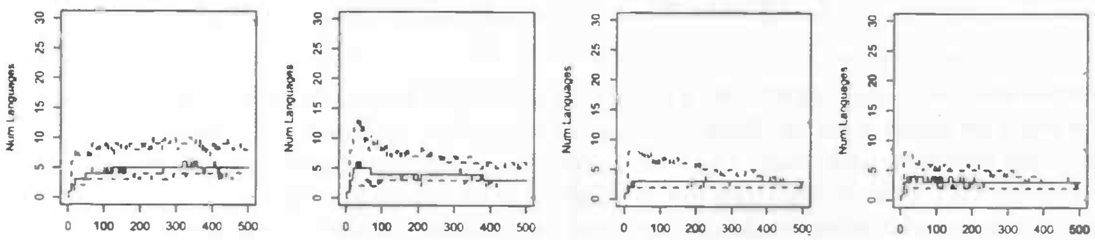
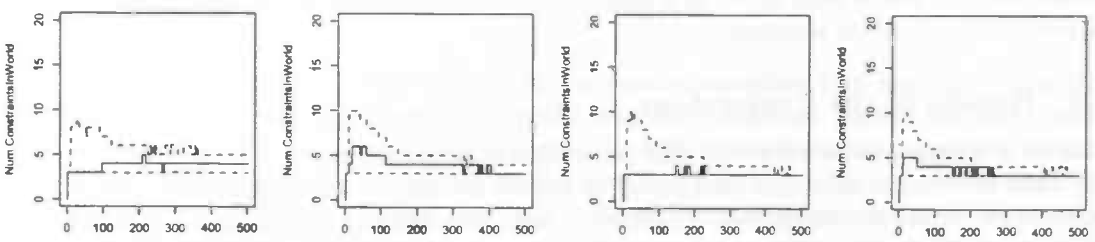


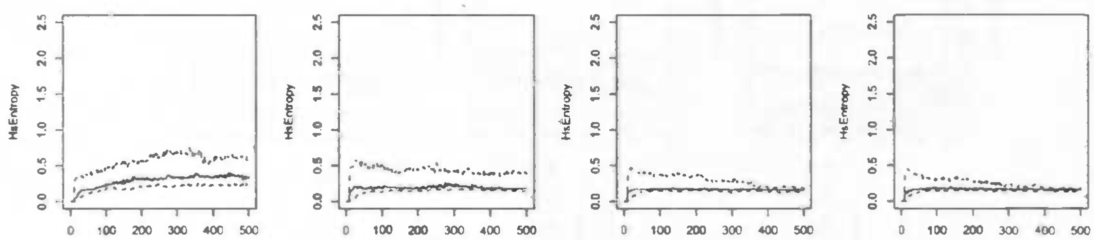
Figure 7.2: Exploration: LanguageVision (0,1,2,3) NumConstraintsInWorld.



In figures 7.4, 7.5 and 7.6 the same graphs are displayed for the LearnFromOthers values 0, 0.25, 0.5, 0.75 and 1.0. LearnFromOthers 0 means that the children are able to learn from others from their birth and 1 means they only learn from their parents.

The number of languages graphs show that a value of 1 causes a trend of continuing growth of the number of languages. This is true because all three quartiles keep growing at 500 time steps. After 500 steps the median languages is approximately 10. The other values cause be-

Figure 7.3: Exploration: LanguageVision (0,1,2,3) HSEntropy. The main difference is visible between the graph of LanguageVision 0 and LanguageVision 2 and 3. Only a textscLanguageVisiob of zero causes a rising trend



tween themselves all similar stabilizing trends with medians around 5. A LEARNFROMOTHERS of 1 causes therefore far more languages in the population. A small difference can be observed between the 3rd quartile of the first and the 2nd to fourth graph, where it lies higher. The NumConstraintsInWorld graphs show that the value of 1 causes an eventually stabilizing trend where the quartiles lie until 370 steps around 3, 7 and 9. After 370 steps the 1st quartile increases to 5. This means there are quite many constraints in the population, which causes the large possibility for more languages to evolve and for larger differences between languages, than when there are for example only 3 constraints. The other graphs show between themselves similar trends in all quartiles, only except for the median in the graph displaying the 0.25 trend which is 3 instead of 4 in the others. The quartiles all have eventually values of 3, 4, 5. The values of the quartiles in the LEARNFROMOTHERS 1 case of 5, 7, 9 case and the other cases of 3, 4, 5 show that, the first causes significantly more constraints in the world. The HSE graphs only emphasize the previous observations. The last graph shows a growing trend in all quartiles, ending with values of approximately 0.45, 0.55 and 1.05 at 500 time steps. The rest of the graphs again show similar trends between themselves; slowly growing with approximate quartile values of 0.2, 0.3 and 0.5. These results indicate a higher diversity.

These observations suggest a higher linguistic diversity when the children only learn from their parents than when they are able to learn from others. It does not matter when they begin learning from others.

In figures 7.7, 7.8 and 7.9 the same graphs are displayed for the PLASTICITY values 0.1, 0.3, 0.5, 0.7 and 0.9. The main observation is that a high PLASTICITY value of 0.9 results in a higher linguistic diversity.

In figure 7.7 there is a main difference visible between the graph displaying the trend of PLASTICITY value 0.9 and the rest of the graphs. In this graph the quartiles stabilize at around 3.5, 5 and 8.5. Between the 0.3, 0.5 and 0.7 graphs there is not much difference and they all display quartiles of approximately 2, 3 and 4. The 0.1 graph even shows lower quartile values of eventually 1, 2, 2. In figure 7.8 the main difference, although not large, is again between the 0.9 graph and the rest. The quartile values in the first stabilize eventually at 3, 4 and 5 constraints. In the 0.3, 0.5 and 0.7 graphs they are 3, 3 and 4 and in the 0.1 graph 3, 3, 3, which is the minimum possible because there are 3 constraints from the beginning. Although there are no huge differences it signifies something if there are 5 in stead of 4 constraints, because then there are more constraint combinations possible and consequently more languages and more difference among the languages. The HSE graphs show and emphasize the difference between the consequence of having a high learnrate of 0.9 and the rest. In this graph the quartiles stabilize at approximately 0.2, 0.3 and 0.7. In the 0.3, 0.5 and 0.7 case they stabilize around 0.15, 0.2 and 0.3. In the 0.1 case these are

Figure 7.4: Exploration: LearnFromOthers (0,0.25,0.5,0.75,1.0) NumLanguages

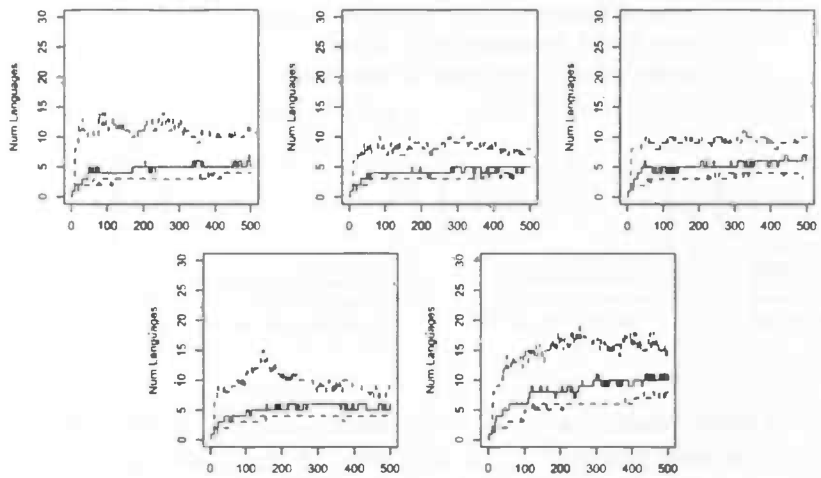
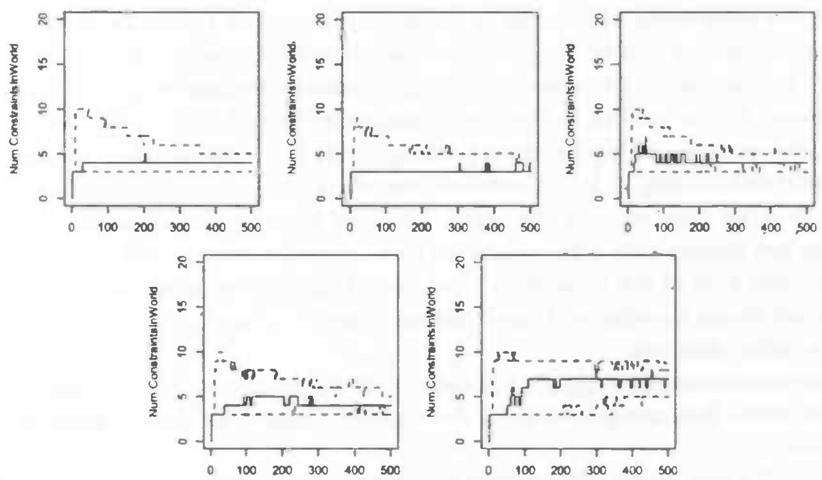


Figure 7.5: Exploration: LearnFromOthers (0,0.25,0.5,0.75,1.0) NumConstraintsInWorld



0, 0.1 and 0.15.

The combination of results conclude in the main observations that first, a PLASTICITY value of 0.9 results in a higher linguistic diversity compared to the rest and second, that a very low learnrate of 0.1 results in the lowest diversity. The difference between the quartiles is around 1,5, 2 and 4.5 languages with the 0.3, 0.5 and 0.7 cases and even 2.5, 3, 6.5 with the 0.1 case. In the NumConstraintsInWorld case but especially in the HSE case these differences are confirmed.

NUMACTIVECONSTRAINTS is the last parameter that has proved to make a difference when varied, even the largest difference. Three experiments have been performed with 2, 3 and 4 active constraints. From the results it is clear that a higher amount of active constraints results in more languages, more constraints and a higher HSE.

In figure 7.10 the NumLanguages graphs are displayed. When there are only 2 active constraints almost none to 1 extra language evolve over time. The median eventually stabilizes on 1 and the 3rd quartile on 2. With three active constraints there are more languages, with quartile values of around 4, 5 and 8.5. When there are 4 active constraints the trend of the quartiles are

Figure 7.6: Exploration: LearnFromOthers (0,0.25,0.5,0.75,1.0) HSEntropy

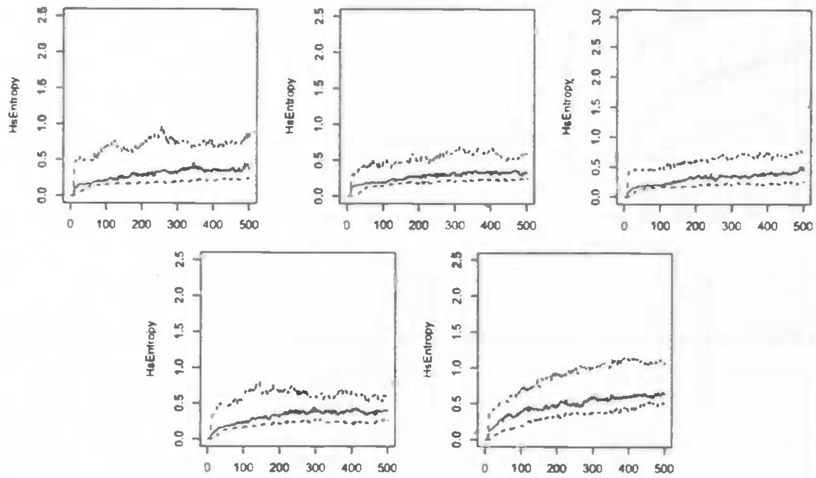
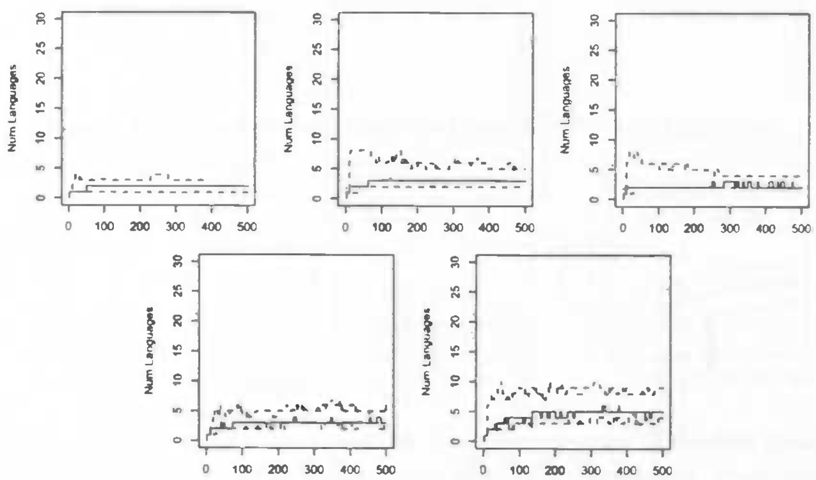


Figure 7.7: Exploration: Plasticity (0.1, 0.3, 0.5, 0.7, 0.9) NumLanguages



all still increasing at the end, with end values of around 15, 23 and 30. The increasing complexity of the language distribution is also reflected in the graphs displaying the NumConstraintsInWorld (figure 7.11). When there are 2 active constraints there are mostly no innovations in the languages, because the median lies steadily at 2 and the 3rd quartile only occasionally reaches 3 constraints meaning a single innovation. With three constraints there are more innovations. The median lies steadily at 4, which means a single innovation. The 3rd quartile eventually stabilizes at 5. With 4 active constraints the trend shows still more innovations, because the 1st quartile is at 5, the 2nd at 6 and the 3rd at seven, respectively entailing 1, 2 and 3 innovations. All the trends show eventual stability in all quartiles. The increasing complexity is very clear in the HSE graphs. With 2 active constraints the 1st quartile and the median lie almost always on 0, where the 3rd quartile is at 0.25. This means the linguistic diversity is almost non-existent most of the times. With 3 active constraints, the 1st and 2nd quartile do have values: 0.2 and 0.3. The 3rd quartile hovers around 0.5 and 0.6. With this fact and the NumLanguages and NumConstraintsInWorld graph it can be deducted there is linguistic diversity. With 4 active constraints the HSE reaches high

Figure 7.8: Exploration: Plasticity (0.1, 0.3, 0.5, 0.7, 0.9) NumConstraintsInWorld

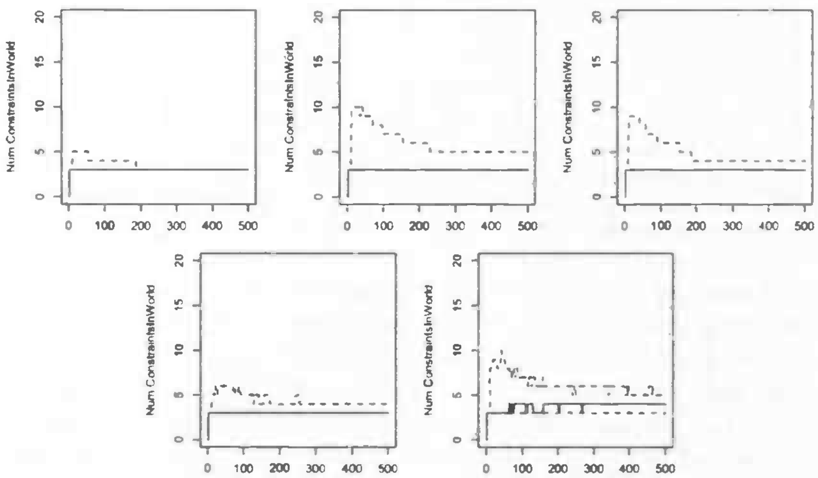
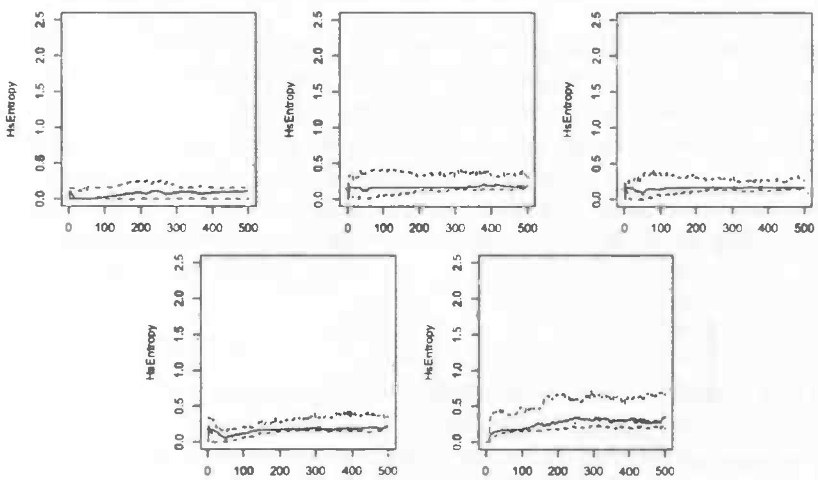


Figure 7.9: Exploration: Plasticity (0.1, 0.3, 0.5, 0.7, 0.9) HSEntropy



values, far higher than the other two, and all quartiles keep on growing within 500 time steps. There is a limit on how many languages there can be (as many as there are adult agents), so the number of languages will stop growing eventually. That there are far more languages is because with 4 constraints there are more languages and therefore more diversity possible.

These results therefore show that 4 active constraints cause a ongoing increase in the complexity of the linguistic diversity within 500 iterations. 3 active constraints cause a moderate and stable complexity and 2 almost no linguistic diversity. The number of active constraints have a clear influence on the linguistic diversity.

Figure 7.10: Exploration: NumActiveConstraints (2, 3, 4) NumLanguages

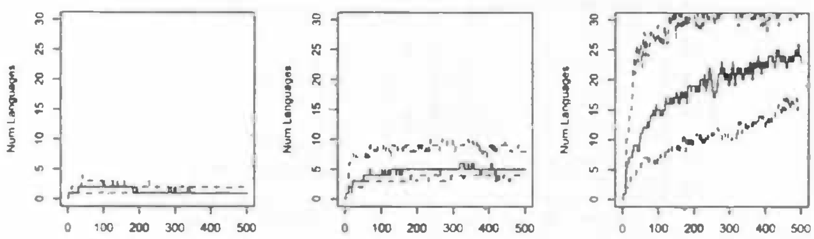


Figure 7.11: Exploration: NumActiveConstraints (2, 3, 4) NumConstraintsInWorld

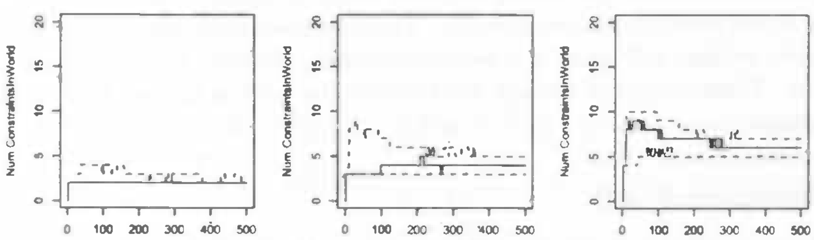
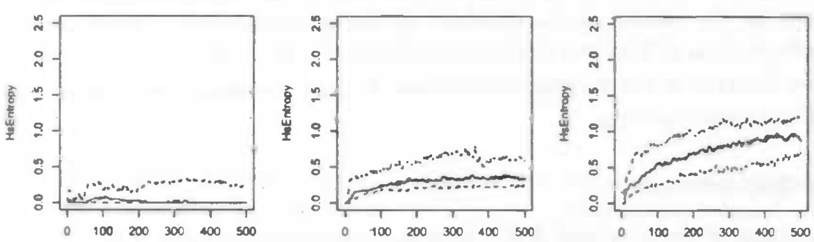
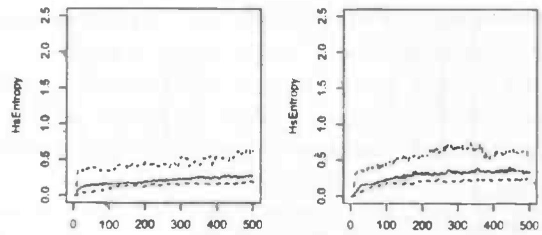


Figure 7.12: Exploration: NumActiveConstraints (2, 3, 4) HSEntropy



The other experiments have produced less interesting results. Therefore these results are summarized and are mostly not accompanied by graphs. The experiment where the grid size is varied resulted in a slightly higher linguistic diversity when it was larger. Surprisingly, although the number of languages is significantly higher when the gridsize is 9x9 in stead of 3x3, the HSE is not much larger. A low ADULTAGE of 1 resulted in a stable trend of more languages and a higher HSE than a high ADULTAGE of 9, therefore in a higher linguistic diversity. Different values of IDEALNUMAGENTS did not seem to have any influence at all. Then there is the experiment where MOBILE is turned of and the LANGUAGEVISION is varied. Here the influence of no movement seems to be non existent. In figure 7.13 there is almost no difference visible. The experiment where the LANGUAGEVISION is 1 and the LEARNFROMOTHERS is varied resulted in similar graphs compared with the graphs where the LANGUAGEVISION has its default value of 0.

Figure 7.13: Not mobile vs mobile with LanguageVision 0 (HSEntropy)



In sum, the results show that the highest LanguageVision, the lowest LearnFromOthers, the lowest Plasticity, the lowest number of NumActiveConstraints, the smallest grid, and the highest AdultAge cause the lowest linguistic diversity within each experiment. The opposite parameter values cause the highest diversity. The hypothesis from these results is that these combined parameter settings will result in a very low linguistic diversity and the opposite settings in a very high one. These combined settings will therefore be used as the two basic experiments in the sensitive study.

7.2.3 The Sensitive Study

First of all, the sensitive study consists of *two basic experiments.* The first basic experiment is with a parameter setting where the linguistic diversity is hypothetically very low and the second basic experiment with a parameter setting where the linguistic diversity is hypothetically very high. These two parameter settings and the hypothesis that they result in very low and high linguistic diversity are based on the results of the previous exploring experiments. After this, a further investigation has been done to find the minimal requirements. First, the parameters setting for a hypothetically low diversity is taken, where every time a single parameter gets the value of which it has in the high parameter setting

The Two Basic Experiments

The parameter settings of the Low and High experiment are given in table 7.2. The parameters that are not included in the table have their default values. Besides the exploration results that indicate low diversity with certain parameter values there are multiple other reasons to hypothesize low linguistic diversity with these settings: there is mobility; very little spatial distribution (the grid is 3 by 3); the learnrate and accuracy is small (plasticity is 0.1) which makes the Discontinuity of Language Transmission (DLT) small (tested in 6); the children are from birth enabled to learn from others which suggests a high degree of averaging; this averaging is even more enforced by the high LanguageVision which enables the children to learn from anybody anywhere on the small grid.

Table 7.2: Parameter settings: High and Low

Parameter	Low	High
GridXSize GridYSize	3	9
AdultAge	9	1
Mobile	Yes	No
Plasticity	0.1	1.0
LearnFromOthers	0	1
LanguageVision	2	0
LearnSteps	500	100
NumAgents	20	100

Results of the two basic experiments

Figure 7.14: Sensitive study: High (Left) and Low (Right) (NumAgents)

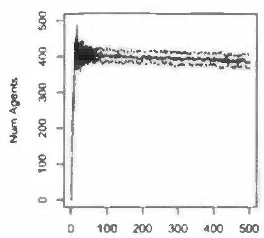


Figure 7.15: Sensitive study: High (Left) and Low (Right) (NumLanguages)

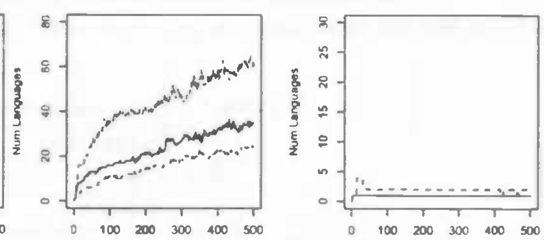


Figure 7.16: Sensitive study: High (Left) and Low (Right) (NumConstraintsInWorld)

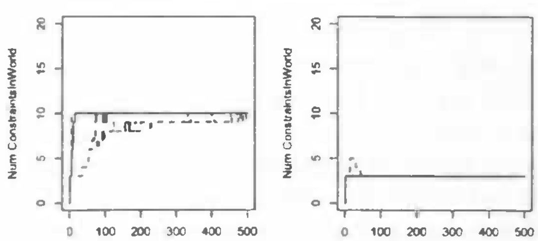
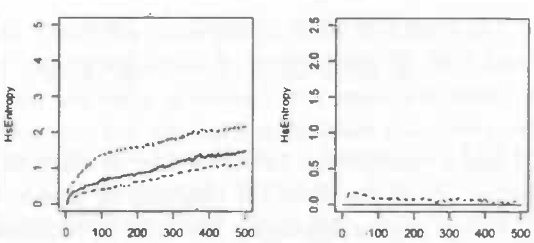


Figure 7.17: Sensitive study: High (Left) and Low (Right) (HSEntropy)



In the figures 7.14, 7.15 and 7.16, the graphs of the high and low parameter settings for the number of languages, the number of constraints and the HSE are displayed. From these graph it becomes clear that the high parameter setting indeed causes a very high and the low parameter setting a very low linguistic diversity. With the high setting the number of languages keeps on growing as time progresses (figure 7.15), where at the end of 500 time steps the quartiles have values of around 20, 30 and 60. Eventually the growth will stop, at a maximum of a single language per adult agent. The NumAgents graphs show the obvious difference between the number of agents in the high and low settings. This is important to show, because the average amount of agents per language indicates if there is sufficient linguistic diversity. In this case the median lies at 30 and there are 400 agents in the population. Because approximately 2/3 of the agents is adult there are only 267 adult agents. A simple calculation results in 8.5 agents per language. Given this, that the number of languages was still growing and that the average number of agents per language should be 20, the linguistic diversity is too high. The latter is emphasized by the trend for the number of constraints in the world and the high and rising trend of the HSE graph. The median and 3rd quartile are as good as always 10, which means that all the constraints are present in the population. The contrast with the graphs of the low parameter settings is high. In the low case the 1st and 2nd quartile of the number of languages graph are 1 and the 3rd quartile remains stable at 2. The number of constraints stabilizes for all quartiles at 3 and the HSE is stable and very low for all quartiles. Having only a single or sometimes two languages is obviously to little linguistic diversity. More is needed to claim that there is real linguistic diversity.

The Exploration with the Low Parameter Setting

The next step is to accurately find the minimum requirements. The idea is to start with the Low parameter setting and perform several experiments where a single parameter has the high value. By slowly working from a supposed low linguistic diversity to a supposed high linguistic diversity, the minimum requirements are probably found along the way. The experiments are shown in table 7.3. The first row means for example that the low parameter setting is used but that the GridXSize and the GridYSize are both 9, in order to test the effect of a larger grid.

Table 7.3: Exploration with Low parameters

Parameter	Low
GridXSize GridYSize	9
AdultAge	1
Mobile	No
Plasticity	1
LearnFromOthers	1
LanguageVision	0

The results of these experiments are partly shown in the first 5 graphs of the figures 7.18, 7.19 and 7.20. In the number of languages graphs the only experiment that resulted in a trend where more languages were present is when the plasticity is put at 1.0. The 1st quartile and the median eventually stabilize at 2 and the 3rd quartile at 3, where the other graphs display quartiles of 1, 1 and 2 respectively. In the number of constraints in the world graphs the same holds in the beginning, but at the finish the quartiles all become 3 just as the others. So not much difference here. In the graphs displaying the HSE it becomes clear that the only difference in linguistic diversity is caused by a high PLASTICITY. The other values do not make a difference at this stage of the exploration.

The idea sprung into mind to test whether the gridsize is of influence with the Low parameter settings and the PLASTICITY put ate 1. In the 6th graphs of the figures 7.18, 7.19 and 7.20, the results of a large gridsize of 9x9 are shown. These show that there is no real influence visible. For this reason and because a 3x3 grid has only room for around 60 agents, which is too low for housing more than 2 languages without being too diverse, from now on, the gridsize will be 5x5.

Because the only increase in linguistic diversity is caused by a high PLASTICITY value, a test has been performed to check whether the PLASTICITY also influences the degree of linguistic diversity the other way around. In other words: does a low PLASTICITY value cause a decrease in linguistic diversity when the High parameter values are used. The results of the High parameter

Figure 7.18: Exploration from Low parameter setting (NumLanguages)

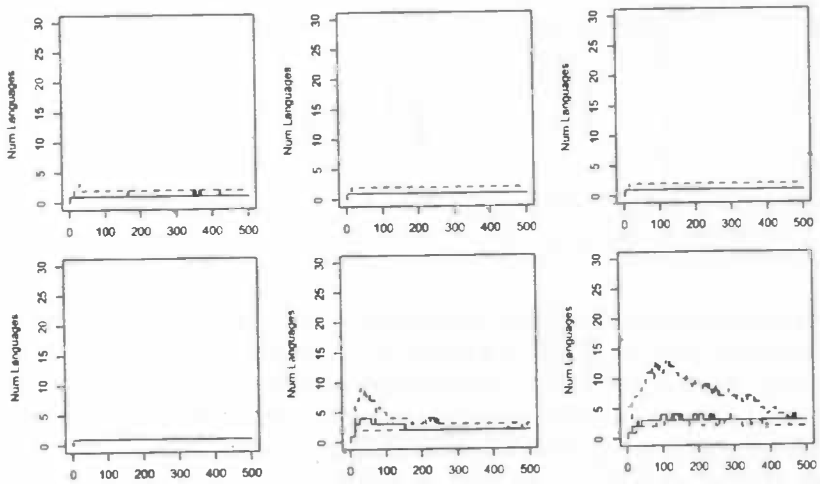
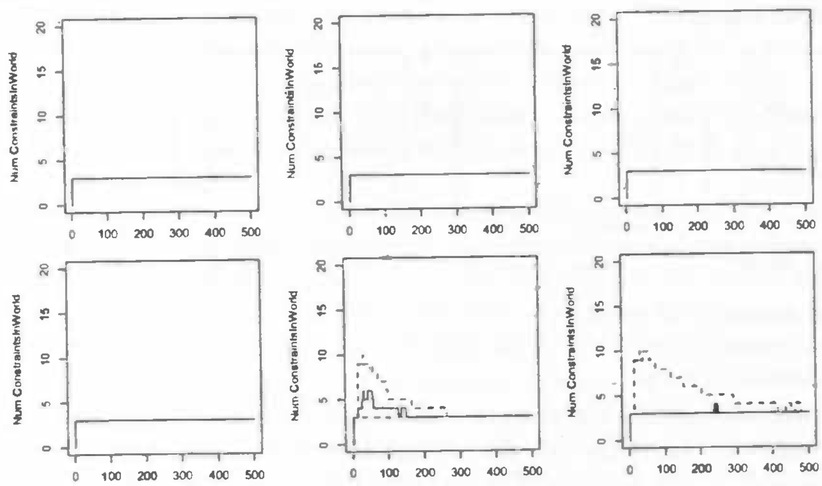


Figure 7.19: Exploration from Low parameter setting (NumConstraintsInWorld)



setting and the High parameter setting and a plasticity of 0.1 are shown in figures 7.21, 7.22 and 7.23. They show clearly that the influence of the PLASTICITY works both ways. The trends of the quartiles of the number of languages lie 20 languages of more lower at the end. Also the number of constraints graphs show a huge difference of approximately 6 to 7 for all quartiles. The HSE entropy confirms the difference, because the quartiles have values of 0.2, 0.3 and 0.45 against 1, 1.2 and 2.

Only a high plasticity value results in a higher linguistic diversity. But the caused linguistic diversity is still too low, with a median of only 2 languages. Consequently, a new set of experiments was performed with the Low parameter setting and the PLASTICITY put at 1 at the base and a high value for another parameter.

Figure 7.20: Exploration from Low parameter setting (HSEntropy)

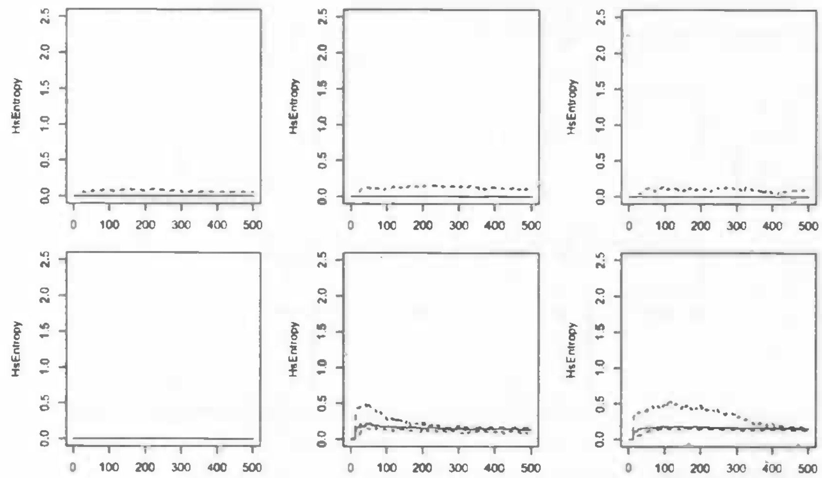


Figure 7.21: High and High with Plasticity 0.1 (NumLanguages)

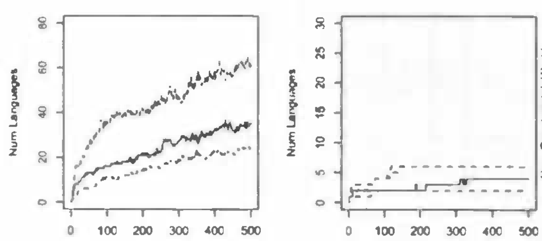
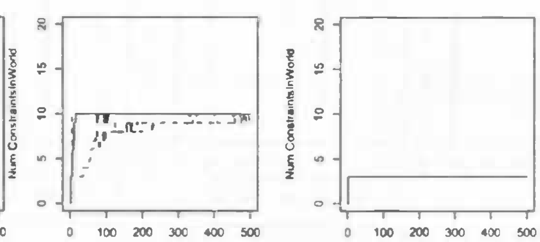


Figure 7.22: High and High with Plasticity 0.1 (NumConstraintsInWorld)



Plasticity 1.0

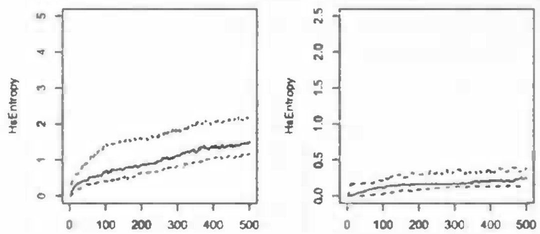
The set of experiments consists of 5 experiments, where subsequently a single parameter is given the High value. See table 7.4. The results show that the high LEARNFROMOTHERS and the immobility

Table 7.4: Exploration with Low parameters and PLASTICITY

Parameter	High
AdultAge	1
Mobile	No
LearnFromOthers	1
NumActiveConstraints	4
LanguageVision	0

of the agents have no increasing influence on the linguistic diversity. The other parameters did make an increasing difference. Therefore, these results are shown in the figures 7.24, 7.25, 7.26. In the NumLanguages graphs the zero LANGUAGEVISION, 4 active constraints and an ADULTAGE of 1, cause higher amounts of languages than with the standard parameter setting displayed in the top-left of the figure. The zero LANGUAGEVISION causes an ever rising median and third

Figure 7.23: High and High with Plasticity 0.1 (HSEntropy)



quartile of languages. The 1st quartile stays stable on 3. 4 Active constraints cause a stable trend of the number of languages, with quartiles of 5, 7 and 8.5. An ADULTAGE of 1 causes a slightly rising trend of all quartiles. Eventually the quartiles have values of 4, 6 and 7. The same higher complexity is reflected in the NumConstraintsInWorld graphs. The zero LANGUAGEVISION causes quartiles of 3, 4 and 5, the 4 active constraints eventually 5, 6 and 6 and the ADULTAGE of 0 eventually 4, 6, 7.5. The HSE graphs give the most complete picture of the results. The standard graph shows quartiles of around 0.09, 0.12 and 0.17. The zero LANGUAGEVISION causes similar trends in the quartiles as the NumLanguages graphs: The 1st quartile stays stable on 0.2, the median rises until a peak of 0.54 and the 3rd quartile rises up till 0.9. The 4 active constraints and the ADULTAGE of 1 cause almost exactly similar graphs with quartiles of 0.2, 0.25 0. 36. These results show therefore that all three parameters cause a slight increase in the HSE.

All in all, these results show that all three tests cause a higher linguistic diversity, because the number of languages, the number of constraints in the world and the HSE show higher trends than those in the standard graphs.

Because the experiments are hypothetically slowly working upwards in linguistic diversity, the current tests seem to be crucial in determining the minimum requirements. This is because the previous experiments all showed too small linguistic diversity, but the current show enough linguistic diversity. It is now time to weigh the current results against the demands for claiming that there is linguistic diversity. These where that there are several groups of more than 20 adult agents with more languages and that the NumLanguages measure, the NumConstraintsInWorld and the HSE all have eventual stabilized trends. In the LANGUAGEVISION 1 case these demands are not all satisfied, because the three measures are not all stable. The NumLanguages and HSE graphs have a still growing trend, after 500 iterations. Nonetheless, the 4 constraints and the ADULTAGE of 1 do satisfy the demands. The 4 constraints cause a median of 7 and a 3rd quartile of 8.5. As mentioned earlier, the number of adult agents is approximately 2/3 of the number of agents. This 160 in this case. This means that there are just enough adult agents per language. The ADULTAGE of 1 cause a median of 6 and a 3rd quartile of 7. This is logically not too many. Also, all the graphs have stable trends, satisfying the second demand.

7.3 Discussion

This is the end of the experiments done in this project. A thorough search has been performed to find the minimum requirements for causing and stabilizing the linguistic diversity. The last experiments with PLASTICITY 1.0, had two good results, where the parameter values are displayed in table 7.5. The first result is with a parameter setting where all parameters have Low values, except for the PLASTICITY and the NUMACTIVECONSTRAINTS, which have values of 1.0 and 4 respectively. The second result is almost the same, except that the NUMACTIVECONSTRAINTS is 3 and the ADULTAGE 1. So, first of all, the minimum requirement is to have a reasonably

Figure 7.24: The NumLanguages graphs for first the standard graph, second with the AdultAge on 1, third with the NumActiveConstraints on 4 and last the LanguageVision on 0 (NumLanguages)

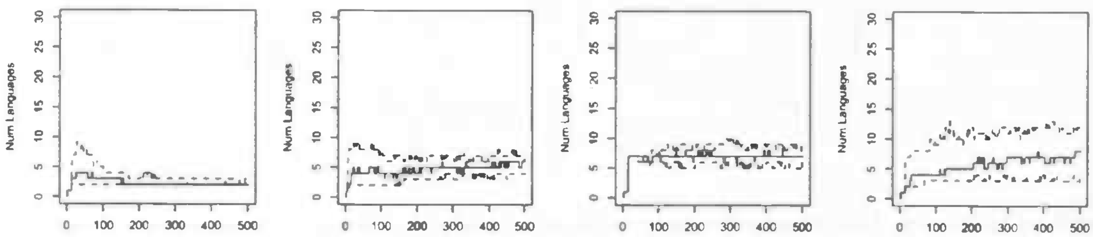
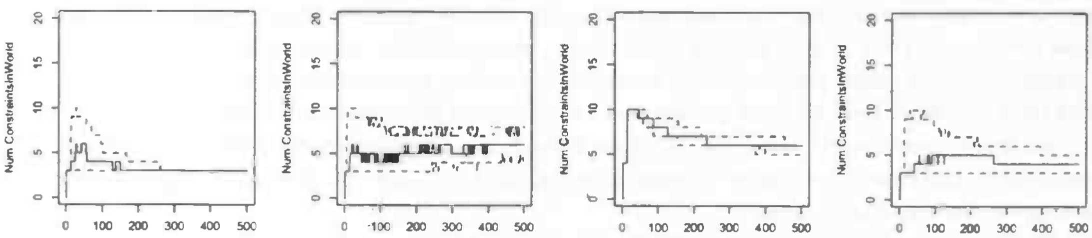


Figure 7.25: The NumConstraintsInWorld graphs for first the standard graph, second with the AdultAge on 1, third with the NumActiveConstraints on 4 and last the LanguageVision on 0 (NumLanguages)



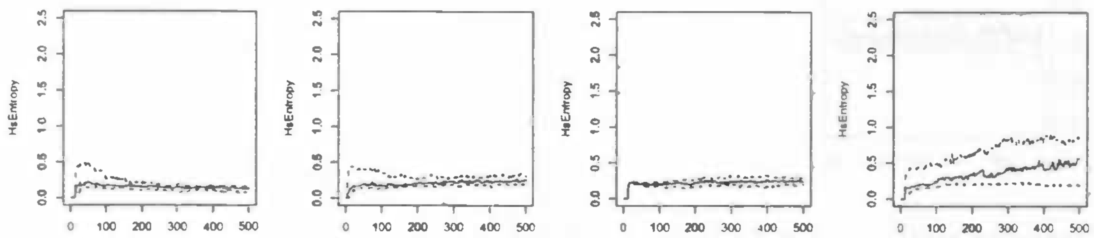
high *Discontinuity of Language Transmission* (DLT), because in both results the Plasticity and therefore the degree of imperfect learning is high and the LanguageScale is average and therefore the imperfect performance average. Then second, either a higher complexity of the language (4 active constraints) or the lowest AdultAge of 1 is necessary.

Table 7.5: Final results

Parameter	Result1	Result2
GridXSize GridYSize	5	5
AdultAge	9	1
Mobile	Yes	Yes
Plasticity	1.0	1.0
LearnFromOthers	0	0
NumActiveConstraints	4	3
LanguageVision	2	2

The model has proven to be capable in testing the several mechanisms of language evolution: the discontinuity of language transmission, the spatial organization and mobility. It is clear that the first mechanism should be high. It is also clear that the mobility can be high. Mobility is a mechanism which slows the evolution of language and therefore a low mobility is not a necessary requirement.

Figure 7.26: The HSE graphs for first the standard graph, second with the AdultAge on 1, third with the NumActiveConstraints on 4 and last the LanguageVision on 0 (NumLanguages)



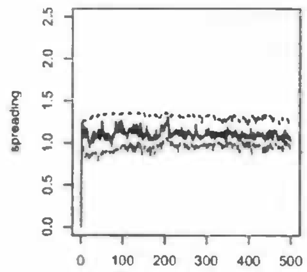
More explanation is needed for the necessary minimum amount of spatial organization, because it is influenced by several parameters. The spatial organization is generally lower when the children are able to learn more from distant individuals. So, first of all, this is influenced by the LanguageVision. Spatial organization is lower when children have a higher radius. In the results the influence became clear in for example the last tests. Another influence is the grid size. If the grid size is high, the spatial organization is higher. The next influence is the LearnFromOthers parameter. If this is high, the children are less able to learn from others and consequently also from distant individuals, causing a higher spatial organization. The next influence is the mobility of the agents. If the agents are mobile, than the children are able to learn from others at other positions, consequently making the spatial organization lower. The last influence is the AdultAge. In the first paragraph of 4.2.2 the influence of it is explained extensively, but it comes down to that a higher AdultAge makes the spatial organization lower, because children can learn from others at other places because they move. Although the influences on the spatial organization of the parameters are dependant of the values of the other parameters (e.g. The higher LanguageVision can only be influential if the LearnFromOthers has not a value of 1, because the child then only learns from their parent), this should give a sufficient picture of what influences the spatial organization.

Now, in both results there is a high LanguageVision, the grid is quite small, the lowest LearnFromOthers age, and a high mobility. This means that the spatial organization for both results is low. For one result the AdultAge is 1, which means the children are only children for the time span of one simulation step, which means the child will not move. But this does not entail it has a much higher spatial organization, because the LanguageVision is 2 and the size of the grid is only 5x5, which means the children have a large part of the total space to learn from others. Nonetheless, this increase in the spatial organization had sufficient influence to increase the linguistic diversity. Remember that the number of LearnSteps is not influenced by the AdultAge.

Throughout the experiments it was learned that, usually, the HSE thus the linguistic diversity kept on growing longer than the number of constraints in the world or the number of languages. Explicit running and viewing several ad hoc experiments showed that over time the set of languages in the world stabilized more and more over time , due to the increasing numbers of agents possessing the languages. This increase in the proportions of agents possessing languages, causes the HSE to rise. In the beginning of a run, new languages emerge, but mostly they vanish soon because there is only a single agent possessing it. Only when the language is spread over more agents it has a larger chance to survive.

In the exploration and in the sensitive study, there where several experiments, where the

Figure 7.27: Typical example of the spreading of languages in the society



LEARNFROMOTHERS was 1. It has to be noted that the development of linguistic diversity with a LEARNFROMOTHERS of 1 could actually be compare with the Iterated Learning Model (IL), because, in every generation, children learn only from a single parent. Although there are multiple adults, there are multiple lines. Because of this relation it is logical that for every such experiment, except for the Low parameter setting experiments, the linguistic diversity kept on rising.

Chapter 8

Conclusion

The goal of this investigation was to answer the following two research questions:

Research question part I What are the minimal requirements for first, causing a group of agents with one language to diverge into several reasonably large groups with more languages, and second, for preserving the achieved linguistic diversity?

Research question part II Will the application of the modern theory of language, Stochastic Optimality Theory, work properly and result into new conclusions?

In this thesis it is shown, by testing the *Individual Language Model* (ILM), that a good representation of language and a mechanism of language transmission capable of producing variation and innovations in the learned language, based on Stochastic Optimality Theory (SOT) and the Maximum Gradual Learning respectively, have been developed. Because the ILM is an abstraction of the essential parts of the SOT and MGLA, which are biological plausible representations and learning mechanisms of language, the ILM is also a biological plausible model. Moreover, the results that came out of the experiments have therefore a certain realism. On the other hand, the abstraction of language is quite a large abstraction. The language only consists of very few features, and many aspects of the MGLA are left out. Moreover, the used representation and mechanism of transmitting language do not deviate much in complexity from the complexity in for example the implementation of cultural features and changes in Axelrod [1997]. Therefore the results and conclusions here, are not so strong as would be in reality and are probably not stronger than in similar investigations. Nonetheless, the results of this project broaden the range of applicability of SOT and maybe even the theoretic range, because this model is an abstraction of and operates over all aspects of language. Moreover, a new (*Spatial Population Model*) that simulates the development of linguistic diversity with, for the first time, a well theoretically based language model has been successfully developed. To answer the second part of the research question literally: The application works properly and has resulted into new conclusions.

This thesis provided an investigation to find the minimum requirements to develop stable linguistic diversity from an uniform language population. It was shown, by the use of and experiments with a *Spatial Population Model* simulating a language population, that the minimum requirements are a high *Discontinuity of Language Transmission* (DLT) and a small *spatial organization* of the population. The DLT is high because the *imperfect performance* is average and the *imperfect learning* is high. The latter requirement needs to be explained a little more, because it actually consists of two independent requirements. These are either that the children have a very low adult age, which increases the spatial organization (although not much), or that the general complexity of language in the population is higher, where the spatial organization is smallest.

That these are the minimum requirements show that it is not necessary to have *geographical isolation*, *social influences*, *functional selection* nor any other mechanisms of language evolution to cause stable linguistic diversity, and that mobility is allowed. Furthermore, the children are allowed to learn from others from the beginning of their lives. To relate this to reality. In 2.3, there was an example of the Black English language diverging from Standard English, despite the day to day interaction with each other. This is remarkable, but the divergence was attributed to the social influences. But the results in this model show that linguistic diversity can emerge without social influences. The high linguistic diversity in New Guinea was due to geographical isolation, but this is also is not needed. Functional selection is a natural internal mechanism which speeds up language evolution, but this model was simplified to leave this out. Therefore, even a internally less capable of changing language is not a necessary requirement in this model. Finally, it is realistic that children learn language mostly from their parents. But this is even not necessary.

Now, it is important to consider the realism of the model and the minimally found requirements in order to justify the previous statements. Earlier, the certain realism of the SOT and MGLA was shown with the scepticism involving the simplicity. In the SPM there are also many simplifications. The children stop learning when they are adult, they cannot learn more languages than one, the grid is discrete, there are no complex social structures, the agents move only with small steps etc.. Although these are simplifications, they are far less invasive than the simplifications done in the language representation and learning. In order to have more realism the representation and mechanism of transmitting language should increase in complexity, thus should be a less simplified version of SOT and MGLA.

What is complex to justify is the high DLT. How does the inaccurate learning of an agent in this model relate to the learning of a child? Furthermore, there is the high percentage of innovations in the language that occur because of this high DLT. Is it realistic to have 8 percent of the language deviate from that of the parent? This seems unlikely. These questions are hard to answer, and it is important to find them out to strengthen or weaken the conclusions drawn here. These are questions to be answered in future work.

8.1 Future Work

The experiments done in this research in order to answer the research questions have been limited. There could be a more sensitive study with the SPM in order to find the minimum requirements more accurately. For example, the moveability in the model was either on or of during the experiments. It would be insightful to investigate the influence of slowing down or speeding up the moveability. Also, throughout the course of the experiments several new measures would have contributed to the insight of the results. Here a lot can be done.

This project has been a starter project, where the SPM has been build from scratch. Very specific research questions have been answered by using this model. Because of this and because the SPM is very flexible and therefore general, it can be used, in its current state, to investigate multiple research questions related to linguistic diversity in general.

But, the model can also be extended in many ways in order to improve the realism. Now, there is for example no multi-lingualism, while this is an important factor in the language evolution. Many people speak several languages or dialects. Because the addition of multi-lingualism would bear significant realism improvement, multi-lingualism would be a worthwhile extension. The language and the language learning mechanism can also be complexioned. Now they are hugely abstracted. The language model could for example be extended with a lexicon, where the words would refer to real things in the world or the mechanism of language transmission could be made more complex and simulate more aspects of the MGLA. Functional selection could be added, in that certain constraints in the language are connected: when one constraint changes the other changes with it. Also, social selection could be added by the selectively choosing of a teacher by a child because of social reasons. Besides these, there are many more extensions imaginable.

8.2 Finally

The implications of this work are hard to place. On the one hand, it has been unknown that stable linguistic diversity was able to evolve with so little influences. No geographical isolation or social influences; who would have guessed that! On the other hand, the realism of the results are a little questionable. Therefore it should be necessary to investigate further and more thoroughly on the subject to make the conclusions even stronger.

Appendix A

.1 Description of the Visualizations of the SPM

The model displays several visualizations to make the development of the simulation more insightful. These consist of graphs which track certain values over time and of spaces which display the spatial distribution of the agents and languages.

In figure 2 a snapshot of the model is displayed. The upper bar, figure 1 contains the main operations like *run* (to run the model continuously), *step* (does a single iteration of the model), *stop* and *refresh* (refreshes the model). At the right the tick count and number of runs is displayed.

There is also a parameter setting window, which is not displayed here. All the parameters are

Figure 1: The upper bar with the STEP, PLAY and STOP and such buttons

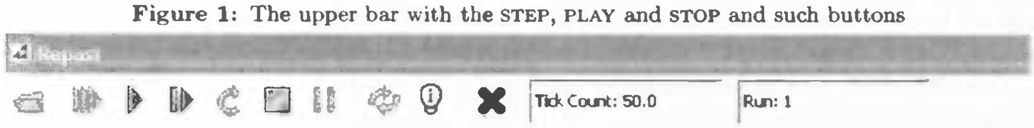
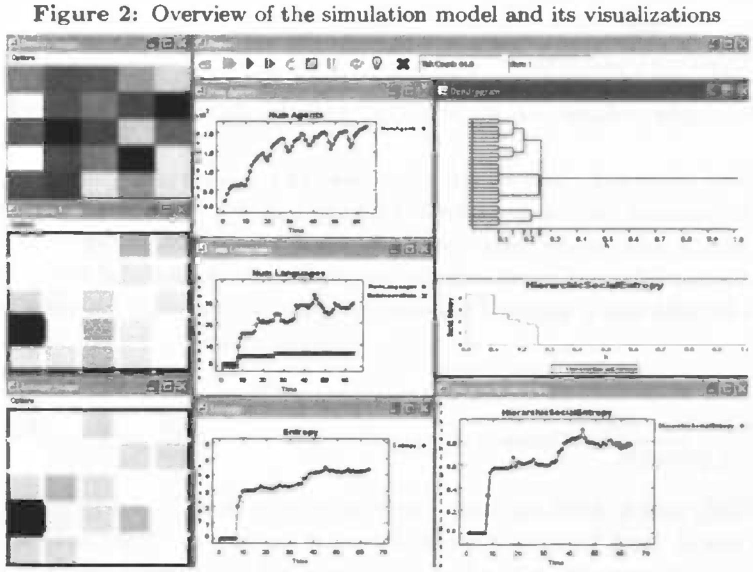


Figure 2: Overview of the simulation model and its visualizations



described in subsection 4.2. Then, there are several views of the scape, which are displayed at the right of figure 2. They display from top to bottom: the density of agents on the sites, the continuous language diversity and the discrete language diversity of the sites and between sites. Four simple graphs track several values of the simulation:

Num Agents This graph tracks the number of agents

Num Languages Tracks the number of languages, the number of constraints in the world and the number of language innovations

Entropy Tracks the simple social entropy of the languages. Higher entropy entails higher language diversity. The entropy measure is explained in subsection 5.1.2

Hierarchic Social Entropy A more complex measure to determine the language diversity. This measure is explained in subsection 5.1.4

The last and most complex figure is the *dendrogram* and is better displayed in figure 6. In this frame two things are displayed: The upper figure displays the dendrogram of the hierarchic social entropy and the lower figure displays the social entropy graph. This graph is explained in 5.1.4.

.1.1 Detailed Explanation of the Several Visualizations

Density Space

The agents occupy a 2D grid and it is convenient to show where the agents are. Because there can be multiple agents on a site, it is difficult to show every agent, especially when there are many agents occupying a single site. Therefore, the density plot is used, which suffices to show the spatial agent distribution, see figure 3, and shows the density of agents at every site at a certain time in the simulation run.

Light red indicates a low and dark red indicates a high density of agents at the site. A white site indicates this is empty. Black indicates the site is overpopulated. A color is determined by the RGB values. The amount of redness is determined by the parameter IDEALNUMAGENTS and the number of agents at the site. This parameter roughly influences the amount of agents in the society, and indicates the average agents per site. If the number of agents equals or is smaller than the IDEALNUMAGENTS the following equation determines the Green and the Blue values:

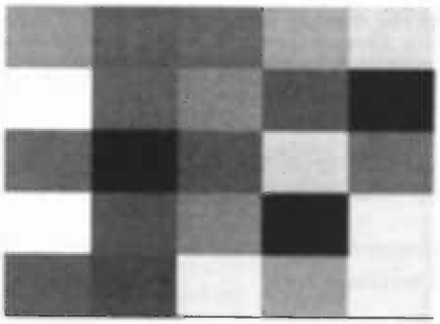
$$Blue = Green = 255 - (NumberOfAgents \cdot \frac{255}{IdealNumAgents}) \qquad (1)$$

The Red value is determined at 255. This way, on an empty site, and an IDEALNUMAGENTS of 10 the RGB color is: Red=255, Green=255, Blue=255, which is white. With 10 agents at the site, this means the RGB color is: Red=255, Green=0, Blue=0. This is an average red.

When the number of agents exceeds the IDEALNUMAGENTS the color has to become redder. With the Red value lower than 255 and a zero Blue and Green value, darker red is created. But with Red values below approximately 160 and lower, the color nears blackness and are not distinctive enough. Therefore the minimum value of redness is MinColorRed=160. This is the equation to determine the Red value:

$$Red = 255 - (NumberOfAgents - IdealNumAgents) \cdot \frac{255 - MinColorRed}{IdealNumAgents} \qquad (2)$$

Figure 3: An example of a density scape. The black sites indicate overcrowding, the white site emptiness, and the site in the second column and 5th row represents a reasonably crowded site



Continuous Diversity Space

Because the subject of this project is language diversity, there is need for a visualization showing the language diversity, see figure 4. This visualization can display areas of high and low diversity and the researcher is able to make estimations.

In this space the local continuous diversity at every site and the continuous difference between languages of adult agents at adjacent sites are displayed. The local continuous diversity is higher when the blueness is higher. The blueness is similarly calculated as the redness of the density sites. Here the amount of blueness is determined by the average continuous diversity between the adults on the site and the maximum difference that two languages can have (MaxDifference). For every combination of two adult agents at the site the average continuous difference between the two languages is calculated by equation 5.1. These are all summed and divided by the number of pairs, which results in a AvgLocalDiversity. The following equation results to calculate the Red and Green value, where the Blue value is 255:

$$Red = Green = 255 - (AvgLocalDiversity \cdot \frac{255}{MaxDifference/2}) \quad (3)$$

With higher AvgLocalDiversity, the Red en Green values are smaller, which results in a darker shade of blue.

With the density scape there was need for a darker red. The same holds for this scape. When the difference is higher than half of MaxDifference, a similar equation is used. With the Blue value lower than 255 and a zero Red and Green value, darker blue shades can be created. But with Blue values below approximately again 160 and lower, the color nears blackness and are not distinctive enough. Therefore the minimum value of blueness is also MinColorBlue=160. This is the equation to determine the Red value:

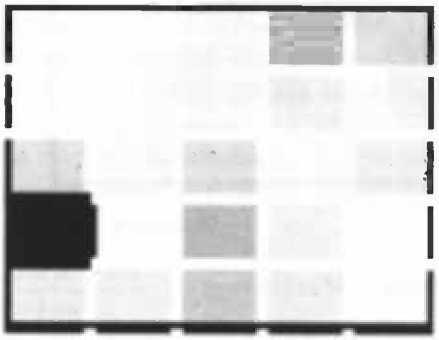
$$Blue = 255 - (AvgLocalDiversity - MaxDifference/2) \cdot \frac{255 - MinColorBlue}{MaxDifference/2} \quad (4)$$

Between sites there are rectangles or borders with a gradient of gray (RGB = x, x, x: The Red, Green and Blue values have equal values to create a shade of gray. Lower values results in darker shades of gray). The grayer the border the higher the difference between sites. Between adjacent sites, the continuous difference between agents of the one site and agents of the other site are calculated and averaged (AvgBetweenDiversity). The continuous difference is again calculated by equation 5.1. The Red, Green and Blue value are calculated by:

$$Red = Green = Blue = 255 - 255 \cdot \frac{AvgBetweenDiversity}{MaxDifference} \tag{5}$$

In figure 4, there are some spots white indicating no diversity and some spots dark blue indicating

Figure 4: An example of a diversity scape. The two white sites in the right bottom corner indicate no continuous diversity and the site in the top row and the 4th column high continuous diversity. Down column three and in between the sites on row 4 and 5, a darker gray indicates a moderate amount of difference.



high diversity. With the density space and the diversity space it is possible to make a cross analysis. For example, low language diversity can be partly explained by a low agent density.

Entropy Space

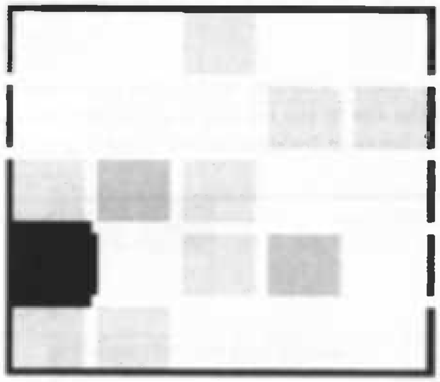
The entropy space is another visualization displaying the language diversity 5. The entropy space equals the continuous diversity space in that a local diversity and the diversity between sites is calculated and shown. The difference is that the local diversity is measured by calculating the *simple social entropy* and the difference between sites with the *discrete difference* explained in subsection 5.1.1. Both measures are discrete measures in contrast to continuous. The continuous difference calculates the average distance between the ranking of the constraints of the two languages (see 5.1.1), where the discrete difference calculates the difference between the order of the constraints.

Simple Social Entropy The entropy measure is based on the number of languages and on the proportions of the languages and is explained in 5.1.2. With a higher number of languages and a more equal distribution of languages over the agents there is a higher entropy.

In this case the simple social entropy is measured for every site which occupies adult agents (This is every site, because a child agent never travels alone). Consequently, for a single site, the proportion(s) of the single or multiple language(s) is/are calculated, and used in the entropy measure. Languages are different when they have a different constraint order. This means this is a discrete measure. In the visualization a high entropy indicates high local diversity, hence a high value of green. The green shades are almost similarly determined as with the blue shades in the continuous diversity scape. There are two differences: first the MaxDifference is replaced by MaxEntropy, second the Red and Blue values are varied in the first equation, 6, and the Green value is varied in the second equation, 7.

$$Red = Green = 255 - (Entropy \cdot \frac{255}{MaxEntropy/2}) \tag{6}$$

Figure 5: An example of an entropy space. Bottom right there are two white sites, where in between the white rectangle indicates no between diversity. On the other hand, the site on the 3rd row and 2nd column is quite green entailing high entropy. Left, right and bottom, the rectangles are quite gray indicating reasonably high in-between diversity with the adjacent and lower sites.



$$Blue = 255 - (Entropy - MaxEntropy/2) \cdot \frac{255 - MinColorGreen}{MaxEntropy/2} \tag{7}$$

The MaxEntropy value is calculated by the following equation:

$$Pi = \frac{1}{NumConstraints!/(NumConstraints - NumActiveConstraints)!} \tag{8a}$$

$$MaxEntropy = log^2 Pi \tag{8b}$$

Just as in the *diversity space* black sites indicate empty sites and white sites entail no diversity.

Discrete Difference The difference between agents at adjacent sites is used here, just as is in the diversity space (see .1.1). The difference is, that the difference between agents is calculated by measuring the discrete difference in stead of the continuous difference.

Discrete vs. Continuous Why are there two displays measuring the diversity in two different ways? First, this is because two different displays cause a more complete view of the language diversity. The advantage of the continuous difference measure is that languages are almost always different. The sensitivity is higher. If the order of the constraints in two languages is the same, making the discrete difference zero, but the constraints are a little different in its ranking, the continuous measure observes a slight difference. In such cases it can be concluded that the constraints are close together, otherwise this small difference is not possible, and that the chances of constraint swapping are high or that the chances of conservation are low.

A drawback of the continuous difference measure is that the order of constraints is neglected. It is easily possible that the continuous difference between languages is small, but the order completely different. The order of constraints is important, because the theory of language in this project is that the order of constraints determines the language and not the ranking. To determine if two languages are different, theoretically, the order of constraints is decisive because these determine the eventual optimal forms, see subsection 3.1.

An advantage of the discrete difference measure is that far less languages are different. This makes the visualization more synoptic, because there is less coloring and language regions are more precisely determined.

It can be concluded that the entropy scape is considered more important in determining the language diversity first, because it displays the theoretically based discrete difference, and second, because it is more synoptic. But, the continuous difference is certainly of importance too.

Discrete Difference and Continuous Difference combined The two displays are also useful because the combination provides additional insight in the nature of the language diversity. Please, look at the fourth upper site from the left in both figure 4 and 5. In the diversity space, this site is very blue, but in the entropy space it is white. This means that the continuous diversity is high but the entropy zero. Further evaluation leads to the observation that the constraints of the languages at this site are quite different in the ranking, but that the order of all languages is the same. This complex observation would not have been possible with one of the scapes. The same can occur with the borders between sites. A general observation is that there can be continuous diversity without a discrete diversity at a site, but not the other way around; there can't be a discrete difference without a continuous difference.

It can be concluded that the two displays can but not necessarily concur in diversity. If the diversity space shows high diversity, it can be that the entropy space displays low diversity. The other way around is less likely but possible too.

Dendrogram

The most complex visualization is shown in figure 6. It encapsulates two figures, the dendrogram and the related plot of the simple social entropy values for every taxonomic level. It provides insight in the hierarchic structure of the language population. A detailed explanation of the hierarchic social entropy is given in subsection 5.1.4

Dendrogram The dendrogram, displayed in the upper part of the figure, is a hierarchic tree representing the language space in the population at the current simulation step, where the previous graphs represented the course of the whole run. At the left, where all the leaves of the tree are, every line represents a single language. More to the right languages are clustered, which indicates that at a maximum difference allowed between subsets of languages (in this case these subsets are single languages) subsets can group into a larger group. The continuation of clustering to the right, indicates larger groups with languages with larger differences eventually ending in a single cluster containing all languages.

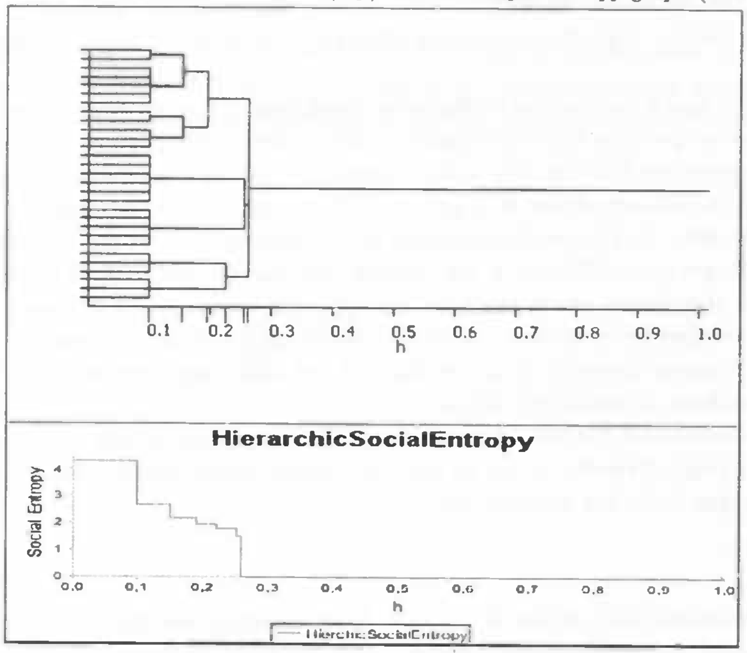
In this figure the levels at which clustering takes place are indicated with a vertical red line on the horizontal axis. These levels represents the so called *taxonomic levels*. These taxonomic levels have a certain h value which represents the maximum difference. They contain clusters with languages.

In the lower part of the figure, the graph represents the simple social entropy values for every taxonomic level. The figure is like a stair going down from right to left. The steps correspond with the clustering of subsets of languages in the upper figure. The *Hierarchic Social Entropy* is calculated by taking the surface size below the graph of the social entropies, see equation 5.4 and equation 5.5.

The *Hierarchic Social Entropy* is introduced because the simple social entropy has several limitations, see subsection 5.1.3, which the HSE can partly overcome. The HSE gives a more complex and complete measure of the language diversity, see 5.1.4 for a detailed explanation.

The dendrogram and the social entropy graph give insight in the language structure. An analysis can be made about the structure in terms of close and loosely related groups of languages, or in more human terms, for example dialects and families of languages. For example,

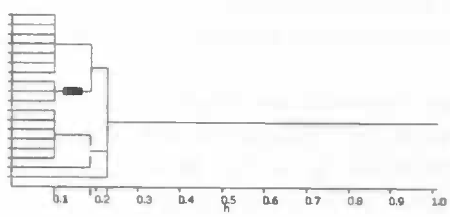
Figure 6: plot of the dendrogram (top) and the simple entropy graph (bottom)



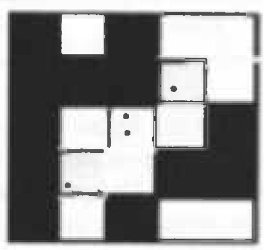
in the dendrogram, 6, the clusters at h 0.1 contain languages which are dialects; the clusters at approximately 2.5, the second red line from the left, contain languages which are distantly related.

Dendrogram and Cluster Selection It is possible to select every cluster with the mouse, as in figure 8(a). The selected cluster has a rectangle to indicate it. By pushing the left mouse button the languages contained in the cluster are displayed in the *Diversity Display* in figure 8(b). Clicking on the clusters therefore provides insight in the spatial distribution of the languages of the selected cluster.

Figure 7: Cluster selection



(a) Selection of a cluster



(b) Languages of this cluster displayed in the language scape

Appendix B

.2 Detailed Description of the ILM Detailed description

The detailed description of the learning system is based on the definition of a language game in ?.

There is a set of *agents* $\mathcal{A}$ of size $\mathcal{N}_{\mathcal{A}}$. There is a single *child-agent* $c \in \mathcal{A}$ and a single or more *adult-agent(s)* $\mathcal{D} \subset A \wedge c \notin D$. Then there is a *lexicon* $\mathcal{L}$ which consists of a fixed set of constraints $\mathcal{C}$ with size $\mathcal{N}_{\mathcal{C}}$, $\mathcal{C} = c_0, ...c_n$ and the associated rankings. Every agent has the same constraints, only the ranking can differ. Every constraint has a ranking value associated with it. Every adult $d \in \mathcal{A}$ has its constraints divided up into a fixed number and set $\mathcal{N}_{\mathcal{P}}$ of *active or participating* constraints $\mathcal{P}$ and a fixed number and set $\mathcal{N}_{\mathcal{I}}$ of *passive or inactive* constraints $\mathcal{I}$. $\mathcal{P} \subseteq \mathcal{C}$ and $\mathcal{I} \subseteq \mathcal{C}$ and

$\mathcal{P} \cup \mathcal{I} = \mathcal{C}$ and

$\mathcal{P} \cap \mathcal{I} = \emptyset$. The *lexicon* $\mathcal{L}_{\mathrm{d}}$ of the adult consists of the active constraints at time t with an associated ranking value $\mathcal{L}_{d,\ t} = \mathcal{P}_{a,t} \times \mathcal{N}_t$. The *lexicon* $\mathcal{L}_{\mathrm{a}}$ of the child consists of all constraints at time t with an associated ranking value $\mathcal{L}_{c,\ t} = \mathcal{C}_{c,t} \times \mathcal{N}_t$. So the adult can be described by a pair $d_t = \langle \mathcal{P}_{d,t}, \mathcal{I}_{d,t} \rangle$. The child has all its constraints initially at zero but has no fixed active or passive constraints during its development. So conveniently the claim is that the child agent has no distinction between active or passive constraints, $c_t = \langle \mathcal{C}_{c,\ t} \rangle$. An agent is generally described by $a_t = \langle \mathcal{C}_{a,\ t} \rangle$

Because the languages change in the process, also the agents change. There is need for a temporal dimension. Agent a at time t is described as $a = \langle \mathcal{C}_{a,t} \rangle$. A time point corresponds to an event where an adult produces an utterance and the child learns from this utterance.

.2.1 The interaction

An interaction $\mathcal{I} =< s, h, u >$ between an adult and a child agent is an interaction between a speaker s and a hearer h. The speaker produces an utterance u, where $u \subseteq \mathcal{C}$. The production of an utterance of an adult proceeds as follows:

1. Every constraint of the adult, including the hidden constraints, are stochastically evaluated, resulting in a temporary constraint ranking $\mathcal{L}_{d,t} = C_{d,t} \times N_t$
2. After evaluation the $\mathcal{N}_{\mathcal{P}}$ highest ranked constraints are determined as the temporary active constraints of the adult resulting in the temporary lexicon $\mathcal{L}_{d,t} = \mathcal{P}_{a,t} \times \mathcal{N}_t$.
3. Out of this temporary lexicon a random set of constraints of size $\mathcal{N}_u$ for which $\mathcal{N}_u \leq N_P$ is produced, resulting in the utterance of the adult at time t $u_t = \mathcal{P}_t \times \mathcal{N}_t$ or $u_t = \langle\langle c_0, ..., c_{\mathcal{N}_u}\rangle, \langle n_0, ..., n_{\mathcal{N}_u}\rangle\rangle$.
4. This set of stochastically evaluated constraints is presented to the child-agent. Here the variable *plasticity* p is introduced. The child will compare the values $\mathcal{N}_t$ of the constraints of the utterance u_t with its own values of the same constraints $\mathcal{C}_{c,t}$. So

$$\forall x_t, n_t(\langle x_t, n_{t,1}\rangle \in u_t \wedge \langle x_t, n_{t_2}\rangle \in \mathcal{C}_{c,t}) \rightarrow ((n_{t,1} > n_{t,2}) \leftrightarrow (\langle x, n_2 + p\rangle \in \mathcal{L}_{c,t+1}) \wedge (n_1 < n_2) \leftrightarrow (\langle x, n_2 - p\rangle \in \mathcal{L}_{c,t+1})$$

This means the rankings of all the constraints which are in the utterance are compared with the rankings in these constraints of the child. If the ranking is higher the ranking-value of the constraint of the child is increased by p, else the ranking-value is decreased by p.

List of Figures

List of Tables

Bibliography

Ash, R. (1999). The sapir-whorf hypothesis. http://www.angelfire.com/journal/worldtour99/sapirwhorf.html.

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution*, 42(2):203–226.

Balch, T. (2000). Hierarchic social entropy: An information theoretic measure of robot group diversity. *Autonomous Robots*, 8(3):209–237.

Bates, E., Thal, D., and Janowsky, J. S. (1992). Early language development and its neural correlates. volume 7, pages 23–32, Amsterdam. Elsevier.

Beecher, M. D., Campbell, S. E., and Stoddard, P. K. (1994). correlation of song-learning and territory establishment strategies in the song sparrow. In *Proceedings of the National Academy of Sciences of the U.S.A*, volume 91, pages 1450–4.

Boersma, P. (1997). How we learn variation, optionality, and probability. In *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, volume 21, pages 43–58.

Boersma, P. and Hayes, B. (2001). Empirical tests of the gradual learning algorithm. In *Linguistic Inquiry*, pages 45–86.

Chambers, J. K. (1983). *Sociolinguistic Theory: Linguistic Variation and Its Social Significance.* Cambridge Press, Cambridge.

Chomsky, N. (1995). *Syntactic Structures.* Mouton & co, The Hague.

Collier, N., Howe, T., Najlis, R., North, M., and Vos, J. R. (2003). Recursive porus agent simulation toolkit. http://repast.sourceforge.net/.

Damerau, F. (1964). A technique for computer detection and correction of spelling errors. *Information Retrieval.*

de Boer, B. (1999). *Self-Organization in Vowel Systems.* PhD thesis, Vrije Universiteit Brussel AI-lab.

de Boysson-Bardies, B. (1999). *How language Comes to Children.* The MIT Press, Cambridge, Massachusetts.

Disner, S. F. (1984). Insights on vowel spacing. In Maddieson, I., editor, *Patterns of Sounds*, pages 136–55, Cambridge. Cambridge University Press.

Elliot, R. (1995). The pinker language instinct: Gradualistic natural selection is not a good enough explanation. University of Pecs, Hungary.

Gordon, R. G., editor (2005). *Ethnologue: Languages of the World.* SIL International, Dallas, Tex, fifteenth edition. Online version: http://www.ethnologue.com.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution.* Cambridge Press, Cambridge.

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

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. *Developmental Neurocognition*, pages 259–274.

Labov, W. (1963). The social motivation of a sound change. In *Word*, volume 19, pages 273–309.

Lanza, E. (1988). Language strategies in the home: linguistic input and infant bilingualism. In Holmen, A., Hansen, E., Gimbel, J., and Jorgensen, J. N., editors, *Bilingualism and the Individual*, Clevedon and Philadelphia. Multilingual Matters Ltd.

LePage, R. B. (1963). Problems of description in multilingual communities. In *Transactions of the Philosophical Society*, pages 189–212.

Livingstone, D. and Fyfe, C. (1999). Diversity in learned communication. In Dautenhahn, K. and Nehaniv, C., editors, *AISB '99 Symposium on Imitation in Animals and Artifacts*, Edinburgh.

Mandelbaum, D. G. (1949). *Selected Writings of Edward Sapir in Language, Culture, and Personality.* University of California Press, Berkeley.

McLaughin, B. (1981). Differences and similarities between first- and second-language learning. *Annals of the New York Academy of Sciences*, 379:23–32.

McLaughlin, B. (1981). Differences and similarities between first- and second-language learning. 379:23–32.

Meillet, A. (1926). *Linguistique Historique et Linguistique Générale.* Honor et Champion, Paris.

Nettle, D. (1998). Explaining global patterns of language. *Journal of Anthropological Archaeology*, 17:354–374.

Nettle, D. (1999). *Linguistic Diversity.* Oxford University Press, Oxford.

Pearson, B. Z., Fernandez, S. C., Lewedeg, V., and Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, 18:41–58.

Pinker, S. (1994). *The Language Instinct.* Penguin Books, London.

Prince, A. and Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar.* Blackwell Publishers, Rutgers University and University of Colorado at Boulder.

Prince, A. and Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science*, 275:1604–1610.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Romaine, S. (1989). *Bilingualism.* Basil Blackwell, Oxford.

Saussure, F. D. (1987). *Course de linguistique générale, édition préparée par Tullio de Mauro*. Payot, Paris.

Sethuraman, N. (1996). Infant language project (theoretical background). `http://cet.middlebury.edu/prgm_midd_baby.php`.

Shannon, C. E. (1949). *The Mathematical Theory of Communication*. University of Illinois Press.

Steels, L. (1998). *Approaches to the Evolution of Language: social and cognitive bases*, chapter Synthesizing the Origins of Language and Meaning Using Co-Evolution, Self-Organization and Level Formation. Cambridge University Press, Cambridge, UK.

Steels, L. and Kaplan, F. (1998). Stochasticity as a source of innovation in language games. In Adami, C., Belew, R., Kitano, H., and Taylor, editors, *Proceedings of Artificial Life VI*, Los Angeles. MIT Press.

Steels, L. and McIntyre, A. (1997). Spatially distributed naming games. *Advances in Complex Systems*, 1(4):301–324.

Trudgill, P. (2000). *Sociolinguistics*. Penguin Books, University of Oxford, fourth edition edition.

Wang, W. S.-Y. and Minett, J. W. (2005). The invasion of language: emergence, change and death. *TRENDS in Ecology and evolution*, 20(5):263–269.

Werker, J. F. and Tees, R. C. (1992). The organization and reorganization of human speech perception. *Annual Review of Neuroscience*, 15:377–402.

Wikipedia (2007). *Co-evolution*. Wikipedia, The Free Encycopedia.