Universiteit Antwerpen
Faculteit Letteren en Wijsbegeerte
Departement Taalkunde

Analogy and Multi-Level Selection in the Formation of a Case Grammar

A Case Study in Fluid Construction Grammar

[Analogie en Multi-Level Selectie in de Ontwikkeling van een Casusgrammatica:
Een Casestudie in Fluid Construction Grammar]

Proefschrift voorgelegd tot het behalen van de graad van doctor in de taal- en letterkunde aan de Universiteit Antwerpen te verdedigen door

Remi VAN TRIJP

Promotoren:
Walter Daelemans
Luc Steels

Antwerpen, 2008

Published by

Universiteit Antwerpen (University of Antwerp)
Departement Taalkunde (Department of Linguistics)
Faculteit Letteren en Wijsbegeerte (Faculty of Arts)
Prinsstraat 13
2000 Antwerpen
België

Official title and subtitle:

Analogy and Multi-Level Selection in the Formation of a Case Grammar:
A Case Study in Fluid Construction Grammar

Nederlandse titel en ondertitel:

Analogie en Multi-Level Selectie in de Ontwikkeling van een Casusgrammatica:
Een Casestudie in Fluid Construction Grammar

Copyright ©2008 Remi van Trijp. All rights reserved.
Cover design: Filip Stoffels

ISBN-978-90-5728-096-2
WETTELIJK DEPOT-D/2008/3677

To Katleen

Acknowledgements

I have always felt that my work of the past few years can best be described as one of “life’s little ironies”. The reason is that in 1866, the *Société de Linguistique de Paris* banned any discussion concerning the origins of language by statute: “La Société n’admet aucune communication concernant, soit l’origine du language, soit la création d’une langue universelle.” And yet it is in Paris that exactly 130 years later, a laboratory was founded that picked up the topic again and shed a whole new light on the matter.

The driving force behind all this is Luc Steels, director and founder of the SONY Computer Science Laboratory Paris and also director and founder of the Artificial Intelligence Laboratory at the Vrije Universiteit Brussel (VUB). He gave me the opportunity to work in both of his labs and involved me in exciting projects such as the *Intensive Science* exhibition celebrating the tenth anniversary of SONY CSL Paris. His ideas and work are the foundations on which I built my own research and I could not have wished for a better mentor. I am also greatly indebted to Walter Daelemans, my supervisor at the University of Antwerp, without whom I would have never been able to embark on my Parisian adventure in the first place. He introduced me to the field of the origins of language six years ago and trained me to become a computational linguist. During the past few years, it was always refreshing and challenging to discuss my work with him and to see things from a different perspective. I would also like to thank him for believing in me, because he (together with Guy De Pauw, who also receives my eternal gratitude) introduced me to Luc and thus got me my first job in science. I would also like to thank the members of my PhD commission and jury for the time and effort they invested in this thesis. It has been a real honour for me to propose my work to such distinguished scientists.

I was fortunate to work in an inspiring environment with lots of symbolic value to me. Sony CSL is located in a side street of *Rue Lhomond*, which gets its name from an 18th century French grammarian. In a city where even a linguist can have a street named after him, I figured, nothing could go wrong. When I first visited Sony CSL I also discovered that I had to take the RER B in the direction of *St. Rémy*, so basically I just had to follow the signs with my name on them in order to reach the lab.

Of course all of this wouldn't have helped a bit if I hadn't been working in a team full of bright people, usually computer scientists who impressed me with their elegant way of thinking and who forced me to look at language in a different way. So I would like to thank my colleagues in Paris (in order of appearance): Martin Loetzsch for showing me the ropes at Sony CSL – such as how to prevent your computer from overheating by putting two of Luc's *Talking Heads* books underneath it – and for helping me out whenever I was chasing a monster in the Babel2 framework; Wouter Van den Broeck for the thought-provoking discussions, the walks through the open-air museum along the river Seine or the Jardin des Plantes, and for always being the first one to believe I was onto something good; and Michael Spranger, a happy-go-lucky Berliner and my Playstation 3-buddy who was capable of soothing my consciousness by saying that playing video games is valid research rather than procrastination. I also wish to express my gratitude to Peter Hanappe and the lab manager Sophie Boucher for their support, and my other colleagues at Sony CSL. I would also like to thank my colleagues from the AI Lab in Brussels, this time age before beauty: Joachim De Beule, the FCG-specialist *par excellence* who challenged me at each step, and who provided me with the technical freedom I so sorely needed; Joris Bleys, who's a very thoughtful person full of surprises (I'm thinking about that monstrous cat in your bedroom!) and who's on his way to become the Mondriaan of computer science; and Pieter Wellens, our team Benjamin whose body seems to be consisting of nothing but shoulders and elbows, and whose enthusiasm, genious and stubbornness will once make a fine mad scientist out of him.

I would be nowhere in life if I hadn't grown up in the most wonderful family one can imagine. I want to thank my parents for working so hard for me, for their love and for giving me all the happiness and opportunities that most people can only dream of. Thanks to my brother and sisters, Bianca, Davy and Tiny; their spouses Paul and Ann; and their children Lorenz, Stef, Laura and my godchild Zoë. Thanks to my family-in-law Flor, Gerda, Griet, Kim, Bart, Joren and Joppe for accepting me in your family and giving me a place to stay each time I needed one. Thanks to all my friends and my 'arch-nemesis' Filip (who made a brilliant cover design for me). Missing all of you has been the hardest part of my life.

Finally, I would like to thank my lovely wife Katleen, for all her love and support; for sticking with me and putting up with my bad moods and stress; and for always being there for me. Seeing your smile each day is more than any man ever deserves, and I couldn't have done this without you. I therefore dedicate this thesis to you. Thank you.

Paris, 22 June 2008

Abstract

Case languages use an inflectional category system for marking event structure. The research in this thesis investigates how such a grammatical system can be developed as the consequence of distributed processes whereby language users continuously shape and reshape their language in locally situated communicative interactions. Since these processes are notoriously difficult to grasp in natural languages, this thesis offers additional evidence from computational simulations in which autonomous artificial agents self-organise a case-like grammar with similar properties as found in case languages such as German, Latin and Turkish.

This thesis hypothesises that language users gradually build their grammar in order to optimise their communicative success and expressiveness while at the same time reducing the cognitive effort needed for semantic interpretation. In the experiments, artificial agents engage in a series of 'language games' in which the speaker has to describe a dynamic event to the hearer. The agents are equipped with diagnostics for autonomously detecting communicative problems, repair strategies for solving these problems, and alignment strategies for coordinating their linguistic inventories with each other. Through comparative simulations, this thesis aims at demonstrating which communicative and external pressures and which cognitive mechanisms are minimally required for the formation of a case grammar.

Two innovating experiments are reported. The first experiment offers the first multi-agent simulations ever that involve polysemous categories. The agents are capable of inventing grammatical markers for indicating event structure and of generalising these markers to semantic roles by performing analogical reasoning over events. Extension by analogy occurs as a side-effect of the need to optimise communicative success and is accompanied by careful abstraction, which yields an increased productivity of the categories. In the second experiment, the agents are capable of combining markers into larger argument structure constructions through pattern formation. The results show that languages become unsystematic if the linguistic inventory is unstructured and contains multiple levels of organisation. This thesis demonstrates that this problem of systematicity can be solved using multi-level selection.

All the experiments are implemented in Fluid Construction Grammar. This thesis presents the first computational formalisation of argument structure in a construction-based approach that works for both production and parsing. It implements the 'fusion' of the participant roles of events with the semantic roles of argument structure constructions. This representation aims at maximal fluidity and introduces some novel concepts in linguistics. Instead of containing a fixed predicate frame, verbs list their 'potential valents' from which the 'actual valency' is selected by argument structure constructions.

Even though the experiments involve the formation of artificial languages, the results are highly relevant for natural language research as well. This thesis therefore engages in an interdisciplinary dialogue with linguistics and contributes to some currently ongoing debates such as the formalisation of argument structure in construction grammar, the organisation of the linguistic inventory, the status of semantic maps and thematic hierarchies and the mechanisms for explaining grammaticalization.

Samenvatting

Talen met een casusgrammatica gebruiken een inflectioneel categorisatiesysteem om de structuur van gebeurtenissen aan te duiden. Dit onderzoek gaat na hoe zo'n grammaticaal systeem ontwikkeld kan worden als het gevolg van gedistribueerde processen waarbij taalgebruikers hun taal voortdurend boetseren tijdens communicatieve interacties. Aangezien deze processen erg moeilijk te vatten zijn voor natuurlijke talen, biedt deze thesis extra data via computersimulaties waarbij een populatie van autonome artificiële agents een grammatica ontwikkelt die gelijkaardige kenmerken vertoont met casustalen als het Duits, het Latijn en het Turks.

De centrale hypothese is dat taalgebruikers hun grammatica geleidelijk aan opbouwen om hun expressiviteit en succes in communicatie te verhogen en om de cognitieve inspanningen te verminderen die ze nodig hebben voor semantische interpretatie. Tijdens de experimenten communiceren de agents met mekaar in 'taalspelen' waarbij de spreker een dynamische gebeurtenis moet beschrijven aan de luisteraar. De agents zijn uitgerust met technieken om communicatieve problemen zelfstandig op te sporen, met herstelstrategieën om die problemen op te lossen, en met strategieën om hun linguïstische kennis op mekaar af te stemmen. Het doel van deze experimenten is om te demonstreren welke cognitieve mechanismen en functionele factoren er minimaal nodig zijn om de ontwikkeling van een casusgrammatica mogelijk te maken.

De experimenten beschrijven twee belangrijke innovaties in de artificiële taalkunde. Een eerste experiment is de eerste multi-agent simulatie ooit waarbij er multifunctionele categorieën gevormd worden. De agents zijn in staat om grammaticale vormen uit te vinden om gebeurtenisstructuren aan te duiden en om deze vormen te generaliseren tot semantische rollen. Extensie door analogie gebeurt als een neveneffect van communicatie en gaat gepaard met voorzichtige abstractie en een verhoging van de productiviteit van de grammaticale categorieën. In een tweede experiment kunnen de agents grotere patronen vormen om argumentstructuur aan te duiden. De resultaten tonen aan dat talen onsystematisch worden als de linguïstische inventaris ongestructureerd is en bestaat uit verschillende niveau's van organisatie. Deze thesis stelt multi-level selectie voor om dit gebrek aan systematiek op te lossen.

De experimenten zijn geïmplementeerd in Fluid Construction Grammar. Deze thesis biedt de eerste computationele representatie van argumentstructuur in een constructie-gebaseerde aanpak die werkt voor zowel productie als interpretatie. Deze representatie implementeert de 'fusie' van participantenrollen met de semantische rollen van constructies. Deze oplossing probeert een maximum aan fluïditeit te bereiken en introduceert enkele nieuwigheden in de taalkunde. In de plaats van een standaard valentiepatroon bevat een werkwoord hier enkel zijn 'potentiële valentie' en wordt de feitelijke valentie bepaald door constructies voor argumentstructuur.

Hoewel de experimenten handelen over artificiële talen, leveren ze ook zeer relevante resultaten op voor taalkunde. Deze thesis gaat daarom een interdisciplinaire dialoog aan met de linguïstiek en levert een bijdrage aan enkele debatten die momenteel gevoerd worden, zoals de representatie van argumentstructuur in constructiegrammatica, de organisatie van linguïstische kennis, de status van semantische maps en thematische hiërarchieën, en mechanismen om grammaticalizatie te verklaren.

"On the one hand, the designers of language have somehow managed to erect magnificent places of sophistication, but for some mysterious reason they failed to clear away the piles of ramshackle irregularities and irrationalities that lie just a stone's throw away. To understand what has brought about this mix of grandeur and folly, we will have to uncover much more of the forces that shape, batter, and renovate linguistic structures."
– Guy Deutscher, *The unfolding of language* –

"Every theoretical description, however exact, turns out to contain errors when you try to implement it."
– Hugo Brandt Corstius, second law of computational linguistics –

"So the reason why I call myself Wonko the Sane is so that people will think I am a fool. That allows me to say what I see when I see it. You can't possibly be a scientist if you mind people thinking that you're a fool."
– Douglas Adams, *So long and thanks for all the fish* –

Contents

List of Figures

List of Tables

Chapter 1

Introduction

It wasn't exactly love at first sight between case marking and me. Or second. I can still recite the complete Latin case paradigm without batting an eyelash because me and my fellow pupils were drilled to do just that: *-us -a -um -i -ae -a*. Like a high-pitched drumbeat of children's voices. Later, at the age of sixteen, I had an unpleasant encounter with what Mark Twain described as *"that awful German language"*. One time you had to say *den* and the other *dem* without any obvious reason. When I asked the teacher about it, I was literally told to just learn the dialogues in the book by rote and trust him that I was saying the right thing. In the meantime, English and French were stealing my heart because they opened new worlds to me without making a fuss about what seemed to be tiny little details at the time.

And yet I found myself dedicating the past three years of my life to the study of case marking. You may interpret this as an unhealthy tendency towards masochism, but I was in fact making amends for my lack of understanding. While working on this thesis, I started to realise that case systems are very elegant solutions to a very complex communicative problem. Case markers turned out to be a grammar's Swiss army knife: they can be used for expressing event structure, spatial and temporal relations, gender and number distinctions, and many other subtle grammatically relevant meanings. I marveled at this unexpected display of functionality and I got intrigued by the rise and fall of case paradigms.

The research in this thesis therefore tries to be a new step in unravelling the secrets of case markers by focusing on the creativity that is involved in their construction. It tries to grasp the living aspects of grammar and how languages are continuously shaped and reshaped by their users. This thesis investigates how case marking systems can be formed as a response to communicative pressures in everyday interactions. But unlike most linguists who gather their evidence from field work, corpora or other studies on natural languages, this research investigates the formation from scratch of a case-like grammar in a population of autonomous artificial 'agents'. In the remainder of this introductory chapter, I will discuss the thesis objectives and its main hypotheses and provide useful pointers for reading this work.

1.1 Thesis objectives and main hypotheses

This thesis is part of a larger research program that tries to move prior work on concept and lexicon formation in populations of artificial embodied agents into the domain of grammar. Its main objective is therefore to reach increased complexity and expressiveness in the artificial languages that are developed by the agents and thus make a contribution to Artificial Intelligence and Artificial Life in general, and artificial language evolution in particular. The second objective of this thesis is to investigate how experiments on artificial language evolution can be relevant for natural language research as well and to engage in an interdisciplinary dialogue with linguistics.

The formation of a case grammar

The general hypothesis underlying this work is that *the development and evolution of grammar is driven by the need to optimise communicative success and expressivity and by the need to reduce the cognitive effort required for processing and semantic interpretation*. This hypothesis is investigated in a case study on the formation of a case grammar for marking event structure in multi-agent computational simulations. The main objective of these simulations is to identify and to demonstrate what the communicative pressures and cognitive mechanisms are that are minimally *required* to allow agents to self-organise a shared case grammar without the need for central control.

Communicative pressures are implemented in the form of **language games**, which are locally situated and routinised communicative interactions. More specifically, the agents play description games in which the speaker has to describe a dynamic event to the hearer. This thesis demonstrates that successful communication about events is possible without grammar, but that grammatical constructions for marking event structure can significantly optimise communication by reducing the cognitive effort needed for interpretation and by avoiding ambiguity.

Cognitive mechanisms are operationalised in the form of computational processes. The goal is to identify which mechanisms are minimally needed to enable the agents to autonomously detect and solve problems during communication, and which alignment strategies the agents need for coordinating their linguistic inventories with each other. This thesis does not explain where these mechanisms come from but rather demonstrates that they are necessary requirements in the formation of a case system by showing the impact of each mechanism on the artificial languages developed by the agents. Special attention is given to two mechanisms: **analogy and multi-level selection**. Analogy is hypothesised to be the main mechanism for innovation which causes multifunctionality of grammatical categories. Multi-level selection is demonstrated to be an indispensable alignment strategy for reaching and maintaining systematicity and hence generalisation accuracy in a language once the linguistic inventory becomes hierarchically structured.

Towards a dialogue with linguistics

The field of artificial language evolution has made rapid progress during the past decade. The advances in robotics and artificial intelligence have made it possible to increase the complexity of experiments and computational simulations, and breakthrough results have been reported with respect to the emergence of vowel systems, concepts, vocabularies, spatial language, colour terms, and grammar.

Researchers in the field have engaged in an interdisciplinary dialogue with computer scientists, biologists, physicists, and people from many other scientific disciplines. Unfortunately, no real dialogue has been held with linguists so far. This is partly due to misconceptions and prejudices about the value of computational modeling and robotic experiments, but mainly because there is no framework yet that enables researchers from both fields to assess each other's work and to put all the pieces of evidence together.

The second main objective of this thesis is therefore to intensify the discussion between artificial language evolution and linguistics, and to look for the conceptual foundations that are necessary for communicating results from one field to the other. In the first place, this thesis wants to offer researchers on artificial language evolution access to the relevant literature in linguistics, which can be considered as the theoretical foundations of this work. On the other hand, it tries to feed the experimental results back into linguistic theory in order to provide linguists with new evidence and alternative hypotheses in some of the currently ongoing debates.

1.2 Navigating this thesis

The research in this thesis is a cross-section between computer science, artificial intelligence and (computational) linguistics. It was therefore written with mainly two kinds of readers in mind: linguists who are interested in the results of computational modeling, and researchers (typically computer scientists) who are already familiar with artificial language evolution or who are interested in performing this kind of research. In this overview, I will therefore mention which chapters are more suited for which kind of reader.

This thesis is divided into three parts. Part I presents the theoretical foundations of this work and sets the scene for the experiments, which are described in Part II. Part III tries to couple the experimental results to some current debates in linguistics and wraps everything together in the final conclusions.

Part I: Theoretical Foundations

Chapter 2 gives more details about the conception of grammar that is underlying the work in this thesis. Linguists will already be familiar with most of the discussions and can read diagonally through the text. For non-linguists or for people who are new to the field, this chapter provides a critical summary of different views in linguistics with respect to the function of grammar and its origins, language change, and the organisation of linguistic knowledge. This chapter also provides the reader with the necessary references to the background literature and offers the reasons for taking a cognitive-functional approach to the formation of case grammar.

The next chapter is especially relevant for linguists and researchers that are not familiar yet with artificial language evolution. It gives a brief overview of the three main approaches in the field based on genetic evolution, reanalysis during child language acquisition, and problem-solving activities in peer-to-peer communication. Each approach is illustrated with respect to the debate on holistic versus compositional languages. The chapter also clarifies the methodology of the research program that is supporting this thesis and gives a brief history of its prior research efforts.

Chapter 4 offers a critical summary that illustrates how linguistic theories are struggling with the problem of argument realisation, a problem which is directly relevant to case grammars for indicating event structure. In this chapter, I also discuss some of the developmental stages in the lifecycle of case markers. This discussion can be read as a general roadmap for investigating the formation of a case grammar. Most linguists will already be familiar with the theories and data presented in this chapter. For non-linguists, however, this chapter again provides more background information and pointers to the relevant literature. Moreover, it aims at demonstrating the enormous complexity involved in grammatical phenomena in natural languages. A good appreciation of this complexity is necessary for moving artificial language evolution forward towards more relevant models since underestimation of the intricacies of grammar have led to misinterpretations and oversimplifications of experiments in the past.

Part II: Formalisation and Experiments

Part II is the most important component of this thesis and is necessarily much more technical than Part I. Diagrams and summaries are provided for allowing people with a less technical background to get the general gist of these chapters. Chapter 5, which is a revised version of van Trijp (2008a), kicks off with a detailed example of how argument structure constructions are formalised in this thesis. This formalisation offers the first computational implementation of argument structure in a construction-based approach. It operationalises the ‘fusion’ of lexical entries with argument structure constructions and introduces some novel concepts to linguistics. The chapter also gives an introductory sketch of Fluid Construction Grammar (FCG) and it places FCG in the broader tradition of computational linguistics.

Chapter 6 gives a detailed overview of the experimental set-up that is used by all the experiments reported in this thesis and makes all the assumptions, abstractions and scaffolds explicit for providing the reader with a clear understanding of what is built in as opposed to what is constructed by the agents. The chapter also reports a replication experiment that confirms prior work on case marking in two-agent simulations and scales this experiment up to a multi-agent simulation. The experiments demonstrate how agents can use grammar for optimising their communicative success and reducing the cognitive effort needed for semantic interpretation. The results also show that analogical reasoning is a powerful mechanism for innovation that can create multifunctional grammatical layers as a side-effect of communication.

The experiments in Chapter 7 introduce an additional mechanism for pattern formation which enables the agents to combine case markers into larger argument structure constructions. The simulations demonstrate however that a problem of systematicity occurs if the linguistic inventory is treated as an unstructured list of linguistic items. The chapter also discusses previous work in the field in which this problem occurred but passed by unnoticed. A solution is offered in the form of an alignment strategy based on multi-level selection in which the evolution of each linguistic item has consequences for all systematically related and all competing elements.

Part III: Impact and Conclusions

Chapter 8 assesses the contributions of this thesis to the fields of artificial language evolution and to linguistics. In a first section, this chapter contrasts the work in this thesis with a recent study which investigates whether the transmission of grammar from one generation to the next can explain the emergence of a case grammar without influence from function or communication. The comparison shows that even though this function-independent model is widely adopted in the field, it fails in almost every aspect and it is significantly outperformed by the cognitive-functional model presented in this thesis. This observation is particularly interesting for people working on artificial language evolution since it exposes some fundamental problems with the 'Iterated Learning Model' that have not surfaced before, since it has always avoided more complex grammatical phenomena and the problem of variation, which is inherent to multi-agent populations.

Chapter 8 also tries to couple the results of the experiments of Part II to some currently ongoing debates in linguistic theory. The remaining sections of this chapter are therefore particularly aimed at linguists who are interested in what kind of evidence or new ideas might come from computational modeling. The first debate deals with the formalisation of argument structure in construction grammar. First, a brief overview is given of argument structure in Sign-Based Construction Grammar (SBCG), a formalism which was still in the process of being implemented at the moment this thesis was written. The comparison between SBCG and the representation offered in this thesis shows that even though FCG is used for supporting experiments on *artificial* languages, it can offer interesting and novel design features which could solve some of the problems that other formalisms are struggling with for *natural* languages. Next, I compare the linguistic inventory that arose in the experiments to how the linguistic inventory is conceived in various construction grammar theories. The experiments in this thesis suggest that linguistic theories have underestimated the problem of systematicity so far and that some kind of multi-level selection mechanism should be implemented instead of or on top of the 'instance links' which are currently being proposed.

In the same chapter, I also discuss some potential applications of the methodology for linguistic typology and grammaticalization. An important debate in linguistic typology concerns the status of semantic maps and whether they represent a universal conceptual space or not. Based on the current results, I will argue that semantic maps could have emerged as a side-effect of analogical reasoning and multi-level selection. I will offer a similar hypothesis as an alternative to thematic hierarchies to explain tendencies in argument selection across languages. Finally, I will provide a possible application for grammaticalization theory. I will argue that the traditionally proposed mechanisms of reanalysis and actualization require a complex cognitive model and that a simpler cognitive architecture could explain the same phenomena using redundancy, pattern formation, analogy and multi-level selection.

Finally, Chapter 9 summarises all the work performed and all the insights gained in this thesis and discusses the next steps that can be taken in this line of research.

1.3 Terminological issues

Throughout this thesis, I have tried to use the most widely accepted terminology in the fields of artificial language evolution and linguistics. Needless to say, however, that in some cases there may be terminological confusion especially when both fields overlap with each other. In this section I will give an overview of some important terms used in this thesis and my understanding of them. Other terminology is explained in the relevant sections when needed.

Emergence is generally used in linguistics in two senses: ontogenetically, it refers to the process of child language acquisition; phylogenetically, it refers to the process in which our species made the transition from a prelinguistic stage to the present stage of linguistic capabilities. This double meaning of the word and the intuitive way it is used in the research field of the origins and evolution of language is confusing, so in this thesis I will apply a more strict definition for several aspects that are commonly grouped together under the notion of 'emergence':

- **Origins**: I will use the word 'origins' to refer to all *genetic* processes involved in the genesis of linguistic categories. This may involve natural selection, exaptation and random mutation.

- **Formation**: 'Formation' means the development of a complex phenomenon with all the necessary ingredients already in place. An example from biology is the formation of an ant path: ant paths are formed in a group of ants that walk around randomly. When an ant finds a food source, it will return to its colony leaving a trail of pheromones. When other ants detect this trail, they will start following it. If a sufficient number of ants do this, a path is formed. The formation of the path does not require any innate or genetic sense of a 'path' inside each individual ant, but rather a population of ants and the pheromone activity. Similarly, the experiments in this thesis investigate the *formation of a grammar for case* in the sense that the agents are endowed with all the necessary ingredients for developing a case grammar, but without having a prior notion of such a grammar.

- **Evolution**: 'Evolution' is used as a general term for all development from less to more complex phenomena. This can be either *genetic* evolution (through natural selection and random mutation) or it can be *cultural* evolution (through the transmission of utterances from one language user to the next). The notion of **selection** is also used as a generic term for all systems in which variations (whether they are genetic or cultural) are in competition with each other. I will use the notion of **multi-level selection** to refer to selectionist pressures operating on *groups* of systematically related linguistic items (see Chapter 7).

- **Acquisition**: I will use the term 'acquisition' in this thesis in the general sense of 'learning' a linguistic item. As such it should not be confused with 'child language acquisition', which involves the specific challenges and situations that children are faced with when learning their language. I will specifically talk about 'child language acquisition' if this more specialised sense of acquisition is meant; and use 'acquisition' if only the process of learning or adoption is relevant.

- **Emergence**: Finally, I will use the term 'emergence' in a more strict sense which focuses on open-endedness: a phenomenon only truly 'emerges' if other developmental pathways were also possible. The development of the phenomenon in one or the other direction is dependent on certain conditions. The difference between formation and emergence is thus that in formation, only one developmental pathway is possible whereas in emergence there is always an alternative. Applied to the topic of this thesis, 'emergence' of a grammar for case would involve at least the alternatives of developing a word order grammar or no grammar at all. Given the specific set-up of the experiments in this thesis, the focus is on the 'formation' of a case grammar.

The formation of grammatical markers involves the development of more lexical or specific markers towards more grammaticality. In order to clearly describe what changes occur in the linguistic inventory, I will apply the same distinction between **generalisation** and **abstraction** as used in Daelemans & Van den Bosch (2005):

- **Generalisation**: Generalisation involves the capacity of successfully handling new cases that go beyond previous experiences. Generalisation does not necessarily involve abstraction but can also be achieved through analogy.

- **Abstraction**: Abstraction involves representations which do not use specific data or instances to handle new situations. Abstraction or **schematisation** of specific instances does not necessarily discard the original data, but may also keep a taxonomy link between the data and the higher-level rules.

The main topic of this thesis deals with the formation of a grammar for case. Unfortunately, a lot of confusing terminology occurs in the literature (for an overview, see Haspelmath, to appear). Throughout this thesis, I will use the following notions:

- **Participant roles**: Participant roles are event-specific roles such as the 'pusher' and the 'pushed' in a 'push-event'.

- **Semantic roles**: I use 'semantic roles' for categories such as 'agent', 'patient' and 'goal' which group together at least two semantically related participant roles. In the formalisation of my experiments, semantic roles occur as slots in argument structure constructions (in which they form a semantic frame) and as 'potential valents' (see below) in the lexical entry of a verb. I also use the term 'semantic roles' for referring to notions in other theories that come closest to my definition, such as notional roles, thematic roles, teta-roles, etc.

- **Syntactic roles or grammatical cases**: Syntactic roles or grammatical cases are categories which make even further abstractions than semantic roles do. 'Syntactic role' is a broader term for covering any category which can be dissociated from a specific semantic role such as 'subject' and 'object'. A 'grammatical case' is what is traditionally understood to be cases such as 'nominative', 'accusative' and 'dative'. Syntactic roles and grammatical cases are assumed to be language-specific just like any other linguistic category in this thesis. I will therefore use for example 'the dative' as a prototypical category which is used for referring to similar but language-specific cases such as the Latin dative.

- **Case markers**: I use the term 'case marker' or just 'marker' to refer to the surface form of a category. This surface form typically maps onto a grammatical case, but in early stages of the development of a case marker it may also map onto a semantic role or directly onto a participant role.

- **Potential valents**: 'Potential valent' stands for the potential categorisation of a participant role as a semantic role; and as the potential syntactic role that a verb assigns to its arguments. Verbs are assumed to only incorporate a 'potential' valency rather than the absolute predicate frames found in most other theories.

- **Actual valency**: Actual valency refers to the pattern of valents that is actually realised as a semantic frame or as a syntactic frame. This is typically achieved by more grammaticalized constructions.

Part I

Theoretical Foundations

Chapter 2

A novel conception of grammar

According to legend, the first linguistic discussion that was ever held ended in both of the debaters drawing their swords... While there is probably little truth to this anecdote, it does make for a good story – not in the least because it illustrates the vigourness with which linguists have always settled their disputes. This point is clearly illustrated by the many different definitions of the object of study itself: language. The disagreements get even bigger when the word *grammar* enters the discussion.

Unfortunately, the consequences of the broad range of available interpretations go far beyond a mere 'confusion of tongues': one's conception of grammar has a major impact on how linguistic theory should look like, what kind of research questions are relevant to be asked, and even what kind of empirical evidence may be used and how it can be interpreted. Every study on the formation of grammatical categories should therefore take a clear position with respect to its assumptions and to other approaches in the research field.

This chapter clarifies the position taken in this dissertation through sections which are all centered around one fundamental issue in contemporary linguistics. Each section basically tries to answer a question raised by the issue, which leads to the next one. Section 2.1 opposes two different views on what the primary function is of language, and illustrates the consequences of both answers for linguistic theory. Next, the nature-nurture debate is addressed, followed by a discussion of theories of language change and grammaticalization in section 2.3. This section also presents the idea of language as a complex adaptive system, which will play a fundamental role throughout this thesis. The organisation and nature of grammar are investigated in section 2.4 and section 2.5 gives an overview of three hypotheses which try to explain the genetic basis of grammar. The final section summarises all the positions taken in the previous sections and argues that there is an ongoing shift of attention in linguistics. This chapter can be read diagonally by linguists who are already familiar with the subject and it offers non-linguists a critical summary of the field.

2.1 The function of language and grammar

The first question we can ask ourselves if we want to investigate the formation of grammatical categories is: *why* do we use language? Intuitively speaking, language seems a way of communicating with other people. The dominant linguistic tradition of the past sixty years, however, rejects this intuitive notion. In this section, I will first discuss this Generativist perspective on the function of language and briefly summarise the objections to it. I will then argue that the main function of language is communication and that the formation of grammar is primarily driven in order to respond to communicative challenges.

2.1.1 The generativist perspective

Before the Chomskyan revolution in the late fifties of the previous century, language was considered to be a complex physiological habit-system that can be acquired through repetition (Bloomfield, 1970, p. 37). Linguistic theory was therefore concerned with the external use of language. Since the mind was conceived as a 'black box', the (mainly) behaviourist approaches of the time did not make any claims about the mental representation of language. It is against this background that we can best understand most of the basic assumptions of Generative Grammar.

Chomsky reacted heavily to the behaviourist tradition, which he called a mechanical one. Instead, he proposed a 'mentalist' vision and argued that linguistic theory should focus entirely on a speaker's internal knowledge about her language. As such, linguists should make an abstraction of this linguistic *competence* (also called *I(nternal)-language*) and not care about *performance* (also called *E(xternal)-language*, also see section 2.3.1). This abstraction can be summarised as follows:

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors [...] in applying his knowledge of the language in actual performance. [...] We thus make a fundamental distinction between *competence* (the speaker-hearer's knowledge of his language) and *performance* (the actual use of language in concrete situations). (Chomsky, 1965, p. 3-4)

The external use of language is thus banned from linguistic inquiry and is also not accepted as the empirical base for deriving the properties of linguistic competence: *"In actual fact, it [performance] obviously could not directly reflect competence. A record of natural speech will show numerous false starts, deviations from rules, changes of plan in mid-course, and so on"* (Chomsky, 1965, p. 4). By taking language away from its actual use, it is no surprise that Chomsky also rejects the hypothesis that communication is the main function of language:

> [E]ither we must deprive the notion communication of all significance, or else we must reject the view that the purpose of language is communication [...]. It is difficult to say what the "purpose" of language is, except, perhaps, the expression of thought, a rather empty formulation. The functions of language are various. (Chomsky, 1980, p. 230)

In his recent work, Chomsky has defended his 'language as an expression of thought'-hypothesis more explicitly. Following Tattersall (1998), he argues that the human intellectual capacity emerged suddenly as *"the result of some genetic event that rewired the brain, allowing for the origin of modern language with the rich syntax that provides a multitude of modes of expression of thought, a prerequisite for social development [...]"* (Chomsky, 2005, p. 3; also see section 2.5). Note that Chomsky sees the intellectual capacity – of which language forms an essential part – as a 'prerequisite for social development'. Quoting scientists such as François Jacob and Salvador Luria, he argues that communication is a secondary function of language which developed as a by-product. Moreover, according to Chomsky communicative challenges cannot provide the necessary selective pressure to produce a complex system such as language.

2.1.2 The impact of communication is everywhere

The lack of a central position for communication in generative grammar is highly problematic for many reasons. A first objection is that if the purpose of language is not communication, then there should be no reason for the members of a speech community to converge on the same set of conventions (also see section 2.3.1). As argued in the previous section, generativists therefore accept that communication is one of the functions of language, but many of them still consider it to be a secondary rather than the primary function. In Chomsky's recent work, he acknowledges that communicative needs have an impact on the language capacity: *"If humans could communicate by telepathy, there would be no need for a phonological component, at least for the purposes of communication"* (Chomsky, 1995, p. 221). He then utters the possibility that the language faculty is almost perfect and that communicative requirements *"might turn out to be "extraneous" to it [the computational system of the language faculty], introducing departures from "perfection" that are satisfied in an optimal way"* (p. 221). In his critique on Chomsky's idea of perfection, Jackendoff (1997) paraphrases Chomsky as follows: *"In other words, language could be perfect if only we didn't have to talk"* (p. 19).

A second objection is much more fundamental and has to do with how sentences are organised: the impact of communicative pressures is everywhere. Functional (e.g. Dik, 1997; Halliday, 1994) and cognitive theories of language (e.g. Lakoff, 1987; Langacker, 1987) point to the existence of grammatical strategies for marking information structure (e.g. topicalisation, focus, prosody, etc.), questions (information requests), deixis (for establishing joint points of reference in discourse), etc. In fact, this whole dissertation tries to show that even basic argument constructions are formed and shaped as a response to communicative pressures. For example, word order, subject-verb agreement and case marking are all communicative strategies which enable the speaker to give clues to the hearer about whom did what to whom in an event. Example 2.1 illustrates this point: the hearer knows that *Jack* is the one who pushed *his brother* and not the other way round thanks to word order. However, if the hearer did not know the conventions of English grammar, ambiguity would remain as to who was the pusher and who was being pushed. If communication did not have a significant impact on grammar, then these grammatical markings would be unnecessary because the speaker knows this information already.

(2.1) Jack pushed his brother.

Finally, even though scientifically speaking it may be very useful to make a distinction between 'competence' and 'performance', every linguistic theory about competence should be compatible with what is known about how this linguistic knowledge is put to use. Many scholars (especially from child language acquisition and psycholinguistics) have criticised Chomsky and other generativists for devising a theory of competence which is inconsistent with data about language use (for an overview, see Nuyts, 1992, p. 99 and further). These arguments have usually been discarded as irrelevant, uninteresting from a theoretical point-of-view, or as part of 'performance'. The competence-performance distinction has thus been increasingly used as an arbitrary way of accepting or rejecting data, or as Lakoff puts it:

> The only consistent way in which I can understand his [Chomsky's] use of the term 'performance' is that he takes it to be a wastebasket for all the phenomena that cannot be accomodated [sic, JN] by whatever theory he happens to be maintaining at a given time. [... I]f like Chomsky, one uses the term 'performance' and 'competence' to characterize what kinds of facts you feel a linguistic theory should be responsible for [...], such a decision can have an effect on whether a given fact is a crucial counterexample, or 'merely a matter of performance' which can be brushed under the rug. Chomsky's shifting definitions of performance provide him with a rug big enough to cover the Himalayas. (Lakoff, 1974, p. 153–155; quoted from Nuyts, 1992, p. 105)

2.1.3 The cognitive-functional alternative

The presence of most (if not all) grammatical constructions and marking strategies in language makes much more sense if we assume that the main function of language is communication. This view means that language should not be regarded as an abstract system that can be studied by itself, but rather as *"an instrument of social interaction"* (Dik, 1997, p. 5). In other words, language has to be investigated by taking the complete communicative cycle into account.

Social interaction through language requires at least one speaker and one hearer who engage in a dialogue involving one or more communicative goals and a specific communicative setting (i.e. the real-world environment and the social-cultural background of the language users). If we accept that the main function of language is communication, then it follows naturally that grammar acts as an interface between meaning and form. Given the enormous expressiveness and complexity of meanings observed in natural languages, it seems obvious that grammars of human languages consist of symbolic units (of varying degrees of generalisation, see section 2.4.2) rather than a signaling system (as commonly found in animal communication systems) or a one-to-one mapping between semantics and syntax (as for example 'universal linking rules' or the Projection Principle in x-bar theory). I will substantiate and operationalise this claim further in the rest of this dissertation.

The fact that grammar seems to organise the syntax-semantics interface logically leads to the hypothesis that grammatical categories and constructions emerged in response to communicative pressures. In other words, grammar developed through problem-solving activities (Langacker, 1991a; Steels *et al.*, 2007): language users try to achieve communicative goals given their cognitive capabilities and given the situatedness of their interactions. This problem-solving process is not a rational conscious one as writing a computer program to complete a task (except perhaps for particular problems such as literary writing) but an intuitive one that is seldom accessible to conscious inspection. As I will argue in section 2.3.3, it is neither an individualistic one but a collective problem-solving process in the sense that there is not one central source of innovation in a speech community, but a *group* of language users which each contribute to and converge on the conventions of a population of speakers. In sum, the three major sources of pressures on the development of grammar should be looked for in the communicative goals, the cognitive skills of the speakers, and the environment in which communication takes place:

> Any natural language can be considered as a particular solution to an extremely complex problem. As with any problem, the possible "space" for arriving at viable solutions is constrained by (i) the nature of the problem itself, (ii) the nature of the problem-solver, and (iii) the circumstances in which the problem must be solved. In the case of natural languages, these three factors can be specified as: (i) the establishment of high-level communicative relationships between human beings, (ii) the biological and psychological properties of natural language users, (iii) the settings and circumstances in which languages are used for communicative problems. (Dik, 1997, p. 7)

2.2 Nature versus nurture

The fact that language primarily serves communication does not necessarily exclude the possibility that grammar is innate (as is the case for many animal communication systems). As I will discuss in section 2.5.1, it may well be that the original pressures for the origins of grammar got hardwired in the brain through a Baldwinian process of genetic assimilation (with possibly further extensions through natural selection), and that nowadays these pressures serve more as a rough guideline for evolution rather than have an immediate effect on the structure of the world's languages. The discussion on what is 'given' and what has to be 'learned' in language is known as the 'nature-nurture'-debate, which can be traced back as far as Ancient India and Ancient Greece.

Robins (1967) writes that the ancient Indian schools of thought wondered whether word-meaning mappings were universal or conventional; and that, similarly, the ancient Greeks debated on whether language was man-made or whether it emerged independently of the language user. Plato argued for the latter opinion in his *Cratylus*, whereas Aristotle defended the view that meanings and language emerge conventionally. Both views are exemplary for the bigger nativist-versus-empiricist split, which runs through the history of science in general.

In modern linguistics, the debate was brought to life again by Noam Chomsky's criticism on behaviourism (Chomsky, 1959) and his subsequent work (e.g. Chomsky, 1965, 1986). In his wake the paradigm of Generative Grammar, which argues for strong nativism, has been the dominant one during the past decades. Recently, however, functional (e.g. Dik, 1997; Halliday, 1994) and cognitive theories of language (e.g. Lakoff, 1987; Langacker, 1987) have become more and more influential. These theories try to explain grammar in terms of more general cognitive mechanisms (such as the ability to create internal representations and categories), and functional principles of sentence organisation (such as information structure, Lambrecht, 1994).

The modern debate is less radical than in Ancient times in the sense that there are no (strongly supported) theories that either argue for complete nativism or for extreme empiricism. Most if not all approaches agree that children are not born as blank slates and that at least *something* has to be innate that makes humans talk in a grammatical, systematic way. This 'something' is also clearly absent in other known species.

2.2.1 Universal Grammar

The dominant view of the past sixty years has been that the human language faculty is equipped with a complex set of prewired linguistic structures, also known as 'Universal Grammar' (UG) or the 'Language Acquisition Device' (Chomsky, 1965). The arguments in favour of such innate linguistic knowledge *"have by now become almost a mantra, a sort of preliminary ritual [...]. Yet these arguments are the reason for the existence of generative grammar"* (Jackendoff, 1997, p. 2). Of all the arguments put forward by generativists, most attention has been given to the 'learnability' of grammar (also see Pinker, 1989, for more on the concept of learnability).

An important question related to learnability is whether children receive enough well-formed input and feedback to be able to learn the language of their parents. In what is called the 'Poverty of the Stimulus' argument or 'Plato's problem' (Chomsky, 1986, p. 7), generative linguists claim that this is not the case, so children need Universal Grammar to fill in the blanks. Moreover, children are said to learn their native language amazingly fast so they must be born with prior knowledge of it.

A second question related to learnability is whether or not certain linguistic structures could ever be created using general learning operators. For example, Jackendoff (1997) argues that *"without considerable inital delimitation [of the acquisition search space by Universal Grammar], it is unlikely that any general-purpose learning procedure, no matter how sophisticated, could reliably converge on such curious notions as long-distance reflexives, reduplicative inflexion, and quantifier scope"* (p. 5).

Based on the arguments of learnability, various proposals have been made about the nature of Universal Grammar. One widespread conception is that UG is a "mental organ" (Piattelli-Palmerini, 1980, p. 76) which grows like a hand or an eye. The assumption underlying this metaphor is that once a person reaches adulthood, her grammar has 'grown' into a fixed and homogeneous set of mathematically precise rules. One major difference with other body parts, however, is that they grow into the same shape in all humans whereas languages diverge enormously from each other. Every theory of UG therefore has to deal with what Tomasello (2005) calls the 'linking problem': *"children may be born into anyone of the several thousand different languages – which vary from one another in myriad ways – and so UG must also have a way of 'linking' itself to particular languages"* (p. 184). The linking problem indeed has been one of the major concerns of defenders of UG (also see section 2.3.1). This requirement, as argued in section 2.1.2, also means that in spite of the basic assumptions of their theory, generativists are forced to accept that communication and language use *do* have an impact on linguistic development.

In its most general formulation one could therefore say that Universal Grammar is a mental module uniquely for the domain of language which (a) determines the space of possible grammars, and (b) specifies how children can use their 'sparse input' to converge on the grammar spoken by the people in their environment. To use the words of Jackendoff (1997, p. 5): *"UG could involve either a specified, constrained search space or a specialized, enriched learning procedure, or both"*.

2.2.2 Did Plato really have a problem?

Even though most of the assumptions of generative grammar rest on the Poverty of the Stimulus argument, it has always been the cause of major controversy in the linguistic community. The counterargument has been roughly the same from the start: children *do* get a lot of structured input from their environment.

The debate about the Poverty of the Stimulus can again be seen as a confusion of tongues, or more precisely, two different views on grammar and the function of language. Generative grammar implicitly reduces a child's learning capacities to *"behavioristic learning theory"* (Tomasello, 2005, p. 189). Linguists who look at the problem from a cognitive-functional point-of-view, however, grant the children more sophisticated cognitive capacities and argue that the communicative setting plays a crucial part in language acquisition. To illustrate this, they have gathered empirical evidence suggesting that children receive highly structured linguistic input which is (functionally) fitting to the communicative setting and environment, and which is adapted to the child's growing communicative competence. As Tomasello argues:

> [T]here is no Poverty of the Stimulus if (1) language is conceived as a set of symbolic units for directing the intentional and mental states of others [...]; and (2) children are given credit for possessing the cognitive and pragmatic skills necessary for learning such meaningful linguistic symbols and constructions, including such things as categorization, analogy, statistical learning, competition among structures, and so forth. (Tomasello, 2005, p. 189)

Studies that gather empirical evidence against the argument of the Poverty of the Stimulus are now covering a broad spectrum of grammatical constructions with various degrees of complexity, ranging from diminutives and number (Gillis, 1997, 1998) to argument structure constructions (Goldberg *et al.*, 2004) and relative clauses (Diessel & Tomasello, 2000).

Perhaps as telling as all these empirical studies is the fact that, despite decades of research efforts, not a single grammatical universal (in the strong, generative sense) has been proposed so far that stood the test of cross-linguistic research. Tomasello (2005, p. 186–189) illustrates this point with an example from the Principles and Parameters framework (Chomsky & Lasnik, 1993), perhaps the most explicit theory on Universal Grammar so far. In the P&P framework, the Language Acquisition Device consists of a set of universal principles and a set of (binary) parameters. Children are hypothesised to set these parameters according to the input they receive from their environment. For example, the pro-drop parameter can be set in such a way that it either allows the subject of a sentence to be dropped (e.g. in Italian) or that the subject has to be expressed (e.g. in Dutch). Once this parameter is tested against non-European languages, however, the characteristics of the pro-drop parameter do not hold anymore. Tomasello (2005) concludes (a) that linguists working in the P&P framework cannot agree on a set of parameters, (b) that parameters do not fit the empirical data, and (c) that a set of innate parameters is an implausible biological mechanism.

Tomasello also suggests that perhaps the *"empirical and logical problems [with these parameters] are the reason that many contemporary theories [...] no longer make use of them"* (p. 188). Indeed, an increasing number of scholars have weakened the definition of Universal Grammar to the general formulation I gave in the last paragraph of the previous section, comprising only constraints on the space of possible grammars and a learning procedure. For example, Steven Pinker writes the following in a response to Brian MacWhinney:

> [A]ny learner who correctly induces a function, theory, or grammar must respect prior ('innate') constraints on its hypothesis space; the data alone are not sufficient. This is a logical point which cannot be denied by any theory, nativist, empiricist, behaviourist, connectionist, constructivist, or emergentist [...]. For Chomskyans, they [the constraints] reside in categories, operations, and principles. For MacWhinney, they reside in the cues, items, alternatives pitted in competition, and categories whose absence constitutes 'indirect negative evidence.' Thus 'conservatism, item-based learning, indirect negative evidence, competition, cue construction, and monitoring' are not 'alternatives' to innate constraints on a learner's hypotheses, but claims about what those constraints are. (Pinker, 2005, p. 949–950)

As stated before in section 2.1.3, a theory of grammar should indeed take into account the constraints that are imposed on possible constructions by a speaker's cognitive apparatus. However, such a generalised definition deprives the notion of 'Universal Grammar' of all its meaning so that perhaps the discussion of what is due to 'nature' and what is due to 'nurture' can finally be investigated starting from more theory-neutral terminology (also see section 2.5 for more on this).

2.2.3 Innate grammatical categories do not exist

The studies on child language acquisition reported in the previous section all suggest that the linguistic input observed by children is both qualitatively and quantitatively rich enough so that the language can be learned. This observation in itself of course does not falsify the nativist stance, so it has to be complemented with research from other (sub-)disciplines. One such research field was already mentioned: language typology.

By carefully describing and comparing languages from various language families, typologists can find out how much grammatical categories correspond or differ from each other across languages. Even though there are still many languages unaccounted for, an enormous amount of various language structures have already been attested and charted into the *World Atlas of Language Structures* (Haspelmath *et al.*, 2005).

Under the influence of the generativist obsession with language universals, a lot of cross-linguistic studies have sailed out with questions such as *what are subjects in these languages*, assuming that there *must* be a subject-category in all languages. As argued by Croft (1991, 2001) and Dryer (1997), this leads to attempts in fitting the observed language into classifications that were made based on western languages. Croft (2001) gives an impressive amount of empirical data to show for example that no discovery procedures (in the sense of the distributional method) can be defined that work across languages: a construction used as a diagnostic to identify a certain category in one language may be absent in another. So either the discovery procedure is different (but on what basis can then be argued that it is in fact the same category?) or the category does not exist in the language. This problem does not only occur across languages, but also within a single language. Similarly, Dryer (1997) argues that many typological studies have been looking for categories which have been proposed in generative frameworks, but which are in fact not necessary for explaining language from a functional perspective.

These studies confirm the intuition of many people working in the field who have been faced with an enormous diversity. It also confirms earlier stances taken by typologically-oriented theories of language to *"take languages seriously"* (Dik, 1997, p. 17). More and more prominent linguists now argue that pre-existing grammatical categories do not exist and have started the debate on what this means for comparative research (Haspelmath, 2007).

2.3 Language as an ecology of constructions

In the previous section I argued that grammatical categories are not innate and that they are language-specific. The next question we can then ask ourselves is: if not innate, then where does grammar come from? In this section, I will defend a principle of 'continuity', that is, the same mechanisms that are observed in present-day language evolution and grammaticalization are also hypothesised to be the mechanisms involved in the emergence of grammatical categories. As Bybee elegantly puts it:

> [G]rammar is constantly changing: old grammatical constructions are constantly being replaced by newly-formed constructions. There is every reason to believe that all existing grammar came about in just the way we observe in the documented cases at our disposal, and that we can put to use what we have learned about this process in trying to understand how the human communicative system comes to be structured grammatically. (Bybee, 1998, p. 250–251)

I will first explain the dominant view that language change is driven by child language acquisition. I will then discuss the empirical objections against this view in section 2.3.2. Finally, I will propose that language should be seen as an ecology of constructions which is constantly being shaped and reshaped through language use.

2.3.1 Child-based theories

The idea that the main source of language change can be found in child language acquisition has been prevalent in generative grammar. This should come as no surprise, given the assumption that grammar reaches a static and non-changing state once

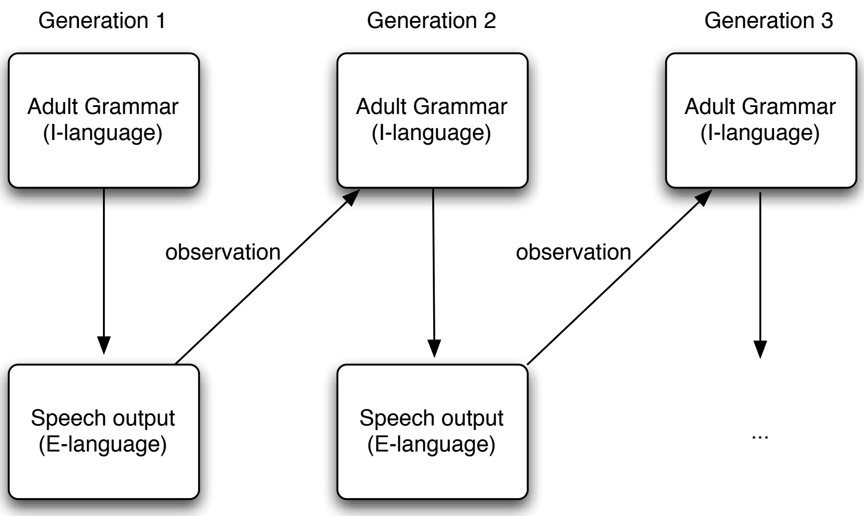


Figure 2.1: In child-based theories of language change, children have to reconstruct the grammars by observing the linguistic output of their parents. Language change is hypothesised to occur when there are mismatches in learning.

a child reaches adulthood. But child-based theories of language change are not exclusive to generative grammar. For example Andersen (1973) incorporates it in his Abduction theory and Hopper & Traugott (1993) (even though they accept changes caused by young adults) present the child-based view as a useful one for conceptualising the mechanisms of language change. Child-based theories of language change are also (sometimes implicitly) assumed by most theories of grammaticalization that regard 'reanalysis' as the most important mechanism in language change.

In the child-based theories of language change, there is no direct link between the grammar of the parents and their children. Rather, children need to 'reconstruct' the grammars of their parents based on the linguistic input they observe. Language change in this model occurs when there is a mismatch in learning between both generations. In the Principles and Parameters framework, this could be a different parameter setting; in some accounts of grammaticalization, this could be (structural) reanalysis. Once the child reaches the adult age, her grammar becomes the new convention and the basis for the next generation of children to reconstruct their grammars. The cycle is illustrated in Figure 2.1 and has been adopted in various small variations since King (1969). The most prominant one of these variations was introduced in Chomsky (1986), in which he makes the distinction between the internal grammars of speakers (called 'I-language') and the external and observable language ('E-language').

2.3.2 Variation among adult speakers

The child-based model is an interesting one because it makes a number of predictions that can be empirically verified. Paraphrasing Croft (2000, p. 45–46; also see Haspelmath, 1998, for similar arguments), these predictions are:

1. The changes to the adult system that are made by children are of the same type as the changes that are attested in language change;
2. The kind of changes made by children are kept during adulthood and thus replace the conventions of the previous generation;
3. Changes are abrupt because they occur within one generation;
4. There is a mismatch between the grammar of the old generation and the grammar of the new one.

As Croft (2000) concludes: *"All four predictions are false"*. I will summmarise his arguments here; for a more detailed investigation and sources see ibid., at 46–53. Croft first surveys a number of comparative studies that match the types of changes found during language acquisition against those found in adult speakers. These studies comprise phonological processes (e.g. word reduction), morphology (e.g. irregular vs regular verb forms), and syntax (e.g. double negation). All studies show that either children produce different kinds of changes than adults; or that when the deviations in child language were the same as those attested in language change, these changes were also made by adults and adolescents.

The second prediction is also countered by empirical data. Field studies show that children tend to converge on the grammars of their parents, so the changes they made during acquisition generally disappear. Croft notes that one of the greatest questions in acquisition research is in fact how these deviations are unlearned. The first 'real' changes or deviations from the adult grammars only occur when children have already grown into adolescence. In other words: after (most of) the acquisition process is already finished.

Historical data and attested examples of language change show that the third prediction made by the child-based theories is also wrong. In acquisition-based changes, new constructions would replace the old ones within a lifetime or within a generation. Many cases of language change, however, can take hundreds of years before the old construction gets lost. In order to account for the gradualness of language change, additional mechanisms such as 'actualization' and 'layering' are needed. I will come back to this issue in section 8.4.3.

Finally, the models predict a mismatch between the grammars of individual speakers in the sense that you either know the 'old' or the 'new' form. This means that although there is variation in a speech community (both forms co-exist because different generations overlap with each other), all speakers would be uniform in their usage. We know from sociolinguistic studies that this is manifestly wrong: language users know several varieties and their "communicative competence" (Dik, 1997, p. 5) allows them to assess which variety is better suited given the pragmatic settings (e.g. formal versus informal speech). This observation is well-captured by Weinreich *et al.* (1968, p. 100), which also reacts against Chomsky's ideal speaker-listener in a homogeneous speech community (see section 2.1.1):

> The key to a rational conception of language change – indeed, of language itself – is the possibility of describing orderly differentiation in a language serving a community. We will argue that nativelike command of heterogeneous structures is not a matter of multidialectalism or "mere" performance, but is part of unilingual linguistic competence. One of the corollaries of our approach is that in language serving a complex (i.e. real) community, it is the *absence* of structured heterogeneity that would be dysfunctional.

2.3.3 Selectionism all the way

I ended the previous section with the observation by Weinreich *et al.* (1968) that language users master several structurally-related varieties of their mother tongue and that speaking a language appropriately also means that you know which variety to select given the circumstances. This view shares my assumption that language primarily serves a social community. I also summarised the arguments put forward by Croft (2000) suggesting that language change occurs among adult speakers rather than during the acquisition process.

Both observations are completely in line with my earlier proposal that building a language can be seen as a problem-solving process in order to reach communicative success (see 2.1.3). In this view, innovations occur in locally situated interactions as a response to specific communicative needs. This innovation then has to propagate in the speech community in order to bring about a change in the language. Variation and innovation are thus common in a population because each individual language user can contribute new innovations. The main challenge of problem-solving models then becomes to explain how innovations may become conventionalised or entrenched in the language, and how individuals can ever agree on a globally shared communication system.

Throughout this dissertation, I will show that a *cultural* selectionist system (as opposed to biological evolution) is the most natural way to explain and model the chain of innovation and propagation of new linguistic items. In this utterance-based selectionist view, language becomes a Complex Adaptive System in its own right, similar to a complex ecosystem (Steels, 2000b). There are many competitions going on: between synonyms for becoming the dominant word in expressing a particular meaning, between different syntactic and semantic categories for certain aspects of the grammar, between various ways in which syntactic categories are expressed, etc. Innovations only survive if they are successful in communication and if they are picked up and further propagated by a sufficient number of language users. Success in propagation can be due to frequency (i.e. the form which is observed most), functional pressures (i.e. a form has found and filled a communicative niche in the language) and social reasons (i.e. a certain variety is also used by speakers with a high social status). The hypothesis that a selectionist system underpins the formation and evolution of language is becoming increasingly popular among linguists. The most extensive evolutionary framework so far has been developed by Croft (2000).

2.4 A usage-based model

The idea that language changes through local innovations and propagation can boast a long history in linguistics, as for example in the work of Ferdinand de Saussure:

> C'est dans la parole que se trouve le germe de tous les changements: chacun d'eux est lancé d'abord par un nombre d'individus avant d'entrer dans l'usage. L'allemand moderne dit: *ich war, wir waren*, tandis que l'ancien allemand, jusqu'au XVIe siècle, conjugait: *ich was, wir waren* (l'anglais dit encore: *I was, we were*). Quelques personnes, influencées par *waren*, ont créé *war* par analogie; c'était un fait de parole; cette forme, souvent répétée, et acceptée par la communauté, est devenu un fait de langue. Mais toutes les innovations de la parole n'ont pas le même succès, et tant qu'elles démeures individuelles, il n'y a pas à en tenir compte, puisque nous étudions la langue; elles ne rentrent dans notre champ d'observation qu'au moment où la collectivité les a accueillis.
>
> All changes germinate in the individual use of language: each one of them is launched first by a number of individuals before entering collective language use. In modern German we say: *ich war, wir waren*, whereas the conjugation in old German up until the XVIth century was: *ich was, wir waren* (in English, we still say *I was, we were*). Some people, influenced by *waren*, created *war* by analogy, which was an act of individual language use; this form, sufficiently repeated and accepted by the community, has become a fact of the language system. But not all individual innovations have the same success, and as long as they reside in individuals, we don't have to take them into account because we study the language system; they only enter our scope of observation from the moment they are welcomed by the collectivity.
> (de Saussure, 1916, p. 138, translation here is my own)

Note that de Saussure makes a distinction between 'la parole' (individual language use) and 'la langue' (the language system). De Saussure is aware that both are mutually dependent because the language system is at the same time the instrument *and* the result of language use. However, as a reaction to the predominant diachronic tradition of his time, de Saussure argues that language use (and its innovations) belong to diachronic studies of language, whereas linguists should focus on the present 'state' of the language system (= synchronic syntax). Even though de Saussure himself was aware that this 'state' can never fully be obtained because languages change over time, the view got nevertheless accepted and even radicalised in many theories of language.

The structuralist revolution unchained by de Saussure (1916) can hardly be underestimated: linguistic theories have tried from then on to describe a certain state of the language system. This practice soon got radicalised into the implicit conception of language as a static and non-changing object, which lies at the origins of for example the child-based view of language change (see the previous section). Even though structuralism had a major impact on all linguistic disciplines, many of its ideas culminated in the generative grammar paradigm. In fact, generative grammar can be seen as an extreme version of structuralism stripped-down of all its references to language as a communication system.

In the next sections, I will first describe the consequences this had in generative grammar on the architecture of a speaker's linguistic knowledge. I will then discuss counterevidence from cognitive linguistics which shows that the generativist model cannot be maintained. Finally, I will argue for a usage-based model of language in which linguists do not have to make an arbitrary distinction between diachronic or synchronic syntax, but in which they can study grammar as an emergent and adaptive phenomenon.

2.4.1 A modular approach

The strategy of making an abstraction of language as a system, leaving all usage-related aspects to historical linguistics, creates the possibility of regarding grammar as a formal, autonomous module. It is exactly this evolution that happened in generative linguistics (and some other theories as well). The reasons for the autonomy of syntax are not far-fetched. Consider the following examples:

(2.2) Colorless green ideas sleep furiously.
(Chomsky, 1957, p. 15)

(2.3) a. Jack pushed the block off the table.
b. The block was pushed off the table.

The much quoted example 2.2 was used by Chomsky to show that sentences can be well-formed syntactically without making any sense, suggesting that syntax is an autonomous system. The sentence pairs in example 2.3 illustrate how the subjects of both sentences map onto different arguments of the 'push'-event in the active and passive constructions. If these sentences are seen from the assumption that language should be studied as an abstract system, one (almost) naturally comes to the conclusion that the notion of a ‘subject’ is entirely formal and does not (directly) relate to the meaning of the sentence. To use the words of Chomsky:

> Despite the undeniable interest and importance of semantic [...] studies of language, they appear to have no direct relevance to the problem of determining or characterizing the set of grammatical utterances. I think that we are forced to conclude that grammar is autonomous and independent of meaning. (Chomsky, 1957, p. 17, quoted from Nuyts, 1992, p. 101)

Functionalists would of course argue that the subject in the sentence pairs of example 2.3 *does* carry meaning, or rather has two functions: it is part of how the event structure is expressed in the sentence and it is used by the speaker for marking information structure (topic versus comment). But the syntax of these two examples seems to be independent of this remark so the principle of the autonomy of syntax remains standing. However, other examples soon emerged in which semantics do have an impact on the ‘well-formedness’ of sentences (also see Chapter 4):

(2.4) I broke his leg. / *I broke him on the leg.

(2.5) I hit his leg. / I hit him on the leg.
(Fillmore, 1970, p. 126, quoted from Levin & Rappaport Hovav, 2005, p. 2)

A purely syntactic theory of grammar would predict that the verbs *break* and *hit*, which both occur in transitive clauses, would have the same syntactic behaviour. Instead, we observe that *hit* can take the complement *on the leg*, whereas this is ungrammatical for the verb *break*. The only possible explanation for these two different syntactic patterns is that the semantics of both verbs is different: *break* involves a change of state in its patient, whereas *hit* involves contact with some surface without necessarily changing its state.

Examples like these have forced generativists to incorporate semantic constraints in their theory and Chomsky’s *Aspects* model therefore included a semantic component (Chomsky, 1965). This did not mean, however, that the autonomy of syntax was given up. Instead, generativists have proposed a modular or componential architecture of grammar with at its heart the (morpho-)syntactic module. Next, there is also a phonological component which gives form to the sentences, and a semantic component. The interface between these components is taken care of by (a small set of universal) ‘linking rules’. Each component also has access to the lexicon, which is conceived as a dictionary-like look-up table of words and idioms.

In this modular approach, syntax is presented as a highly abstract calculus or as a formal system that can generate all well-formed sentences of a language. All semantic constraints are specified in the lexicon, including the valency of verbs (see Chapter 4). When describing a language, the linguist thus has to decide which aspects of the language fall into the syntactic module and which ones belong to the lexicon. In this view, the distinction is made between 'core' grammar and 'periphery': the 'core' contains all the abstract rules whereas the periphery contains all the exceptions to the rule. The distinction between core and periphery is not exclusive to generative grammar, but applies to all linguistic theories that argue for a sharp distinction between grammar and the lexicon.

2.4.2 A continuum of constructions

The distinction between a 'core' and a 'periphery' has led to increasing frustration with linguists. For one thing, classifying empirical data in either of these categories largely rests on theory-internal criteria or arbitrary decisions, which makes it impossible to falsify hypotheses. Moreover, research in natural language processing (NLP) shows that there are an awful lot of 'exceptions' in language. Daelemans & Van den Bosch write the following:

> NLP tasks are hard to model and also hard to learn in terms of broad-coverage rules, because apart from obvious regularities, they also tend to contain many sub-regularities and (pockets) of exceptions. In other words, apart from a core of generalizable regularities, there is a relatively large periphery of irregularities [...]. Daelemans & Van den Bosch (2005, p. 123)

In some cases, like the German plural system, the notion of what is 'regular' and what are 'exceptions' becomes very blurry because the default rule is less frequent than the cases which are supposed to be 'irregular' (ibid., at 27).

Other evidence comes from the study of idiomatic phrases. Traditionally, idioms such as *home is where the heart is* are classified as part of the lexicon. However, language is full of idioms: *wide awake, all of a sudden, to kick the bucket, answer the door,* etc. Influential studies such as Fillmore *et al.* (1988) and Nurnberg *et al.* (1994) show that most 'idioms' cannot be simply dismissed as lexicalised items. For example, *kick the bucket* still requires a subject, agreement and/or tense marking on its verb as in *John kicked the bucket*.

In other words, some aspects of idioms are like lexical items in that they cannot be replaced by anything else and that (at least part of) the meaning is conventional and unpredictable from its parts. On the other hand, some aspects behave like grammatical constructions such as tense marking. In fact, there is a whole continuum of idiomatic constructions, which ranges from more idiomatic to more schematic. A couple of examples are:

(2.6) What's a girl like you doing in a place like this?
(The 'what's X doing Y'-construction, Kay & Fillmore, 1999)

(2.7) The sooner you get started, the sooner you finish.
('the X-er, the Y-er'-construction)

(2.8) They have opened a bar.
(perfective construction: SUBJ + *have* + VERB + *-ed*)

These data suggest that the distinction between the lexicon and grammar is not a sharp or a discrete one. Many cognitive theories of language such as Cognitive Grammar (Langacker, 1987) and Construction Grammar (e.g. Goldberg, 1995; Croft, 2001) in fact grew out of a concern of these more idiomatic constructions. They all suggest that the componential architecture of generative grammar (i.e. lexicon vs syntax vs semantics) cannot be maintained any longer.

2.4.3 Emergent grammar

The insight that there is no sharp distinction between lexical and idomatic items on the one hand, and highly schematic constructions on the other, has posed a serious challenge on linguistic theory: how can rules and exceptions be accounted for in a unified theory? Here again, a natural solution can be found if language is put back into its natural habitat of social interaction. In other words, we need a functional explanation for language, or to use a more fashionable term these days: a usage-based model (Barlow & Kemmer, 2000).

In usage-based models, grammatical constructions are continuously shaped and reshaped as a response to their usage in every-day communication. One fundamental aspect of usage-based models is the objection to what Langacker (2000, p. 2) calls the 'rule-list fallacy' (i.e. the assumption that rules and instances are mutually exclusive). As an alternative, the usage-based approach allows instances and *schemas* (or rules) to co-exist in memory. The more schematised linguistic units are hypothesised to be constructed and acquired on the basis of instances in a stepwise fashion; and language learners are assumed to possess a sufficient amount of general cognitive mechanisms for doing so. The existence of innate knowledge is not completely discarded but only used as a last resort if all other explanations fail.

The co-existence of schemas and instances and the bottom-up fashion in which language is acquired means that usage-based models allow for an enormous amount of redundancy as opposed to the traditional reductionist theories in linguistics. From what is known from psycholinguistic research so far, this is cognitively speaking the most realistic approach. The broad family of usage-based models vary a lot as to how much redundancy they allow, which is closely connected to the status that is given to schemas and instances. On one side of the scale, there are models in which the instances are discarded as soon as a more schematised construction is acquired (e.g. Kay & Fillmore, 1999). On the other side of the scale, exemplar-based models feature *only* instances (e.g. Daelemans & Van den Bosch, 2005; Keuleers & Daelemans, 2007; Skousen, 1989). In the latter models generalisation is achieved through analogical or similarity-based reasoning over stored instances without making abstractions. Most in-between approaches allow for the co-existence of instances and schemas (e.g. Croft, 2001; Langacker, 2000; Verhagen, 2006).

The work in this thesis allows for *careful abstraction*, which groups together families of related instances rather than imposing an abstract, top-down schema which needs to be overridden locally by lower level schemas or instantiations. The main reason for this approach is that the grammatical conventions that arise in the experiments are not fixed from the beginning and still have to be negotiated by the language users, so forgetting instances or making unwarranted abstractions can have disastrous consequences for a speaker's communicative accuracy. Each abstraction therefore needs to keep the full functionality of the instances from which it was constructed. Another reason for only allowing careful abstraction is that forgetting 'exceptions' (by editing them or by abstracting away from them) can be harmful for generalisation accuracy as well. This has been demonstrated for various NLP tasks such as part-of-speech tagging, prepositional-phrase attachment and base noun phrase chunking (Daelemans *et al.*, 1999). Recent work, however, shows that careful abstraction can lead to impressive compression rates without harming generalisation accuracy (Daelemans & Van den Bosch, 2005; van den Bosch, 1999, to appear). I will come back to the approach adopted in this thesis in Chapter 6 when explaining the experiments.

Despite the many differences between various usage-based models, most of them propose two factors that decide whether a linguistic item becomes conventionalised and achieves unit status in the linguistic inventory: frequency and function. Roughly speaking, every form which occurs frequently enough is stored in memory. For example, if the plural form *boys* occurs frequently, the instance gets its own unit status even though it can be derived from a more schematic construction [NOUN + *-s*]. This hypothesis fits the observation that irregular forms are among the most frequent ones in language hence they are able to withstand the competition from a more regular schema in the language. Fully stored instances can be seen as an optimisation strategy because they can be readily used without further processing. When the language user creates an abstraction of instances into a higher level unit, this means the construction of a new, co-existing unit in the inventory rather than the replacement of the instance.

The bottom-up and redundant approach of usage-based models perfectly suits the needs of the experiments in this thesis: it allows for all linguistic items to be treated in a uniform way and it acknowledges the impact of communication on grammar. At each interaction the linguistic inventory can be changed: frequently co-occurring words can be combined into new patterns, specific categories can be extended to cover new situations, etc. This matches the previous statements in this chapter that grammaticalization occurs in language use and that the linguistic inventory may be adapted as the result of problem-solving activities during communication. As such, the usage-based model breaks the radical dichotomy between diachronic and synchronic grammar, which has dominated linguistics for nearly a century now. Instead, grammar should be regarded as an emergent phenomenon (Hopper, 1987) which is constantly on the move and which is constantly adapting itself to new communicative pressures.

2.5 The genetic constraints on language

As I briefly mentioned in section 2.2.2, the notion of 'Universal Grammar' has been generalised so strongly in linguistics that the nature-nurture debate can now be held on the basis of a more theory-neutral question: what are the innate constraints that restrict the space of possible grammars? In this section, I will briefly mention three possible explanations: adaptation, exaptation, and recruitment.

2.5.1 Language as an adaptation

A widespread and popular view since the influential work of Pinker & Bloom (1990) is that language evolved as an adaptation through the process of natural selection. In its essence, the adaptationist view claims that natural languages are too complex (a) to be learned without resorting to prior innate structures, and (b) to be the result of a random mutation or exaptation in genetic evolution (see the following section). Communicative success then serves as the selectionist drive because it influences the fitness of members of the population and thus the spreading of genes. Pinker & Bloom suggest that learning mechanisms existing prior to language could have become innate through the Baldwin effect (Baldwin, 1896), which shows how cultural learning mechanisms can sometimes guide the process of natural selection. These learning mechanisms, however, only provided the raw material that underwent further processes of adaptation through natural selection.

In a more recent paper, Pinker (2003) considers proposals that argue that language is not an adaptation itself, but the result of more general cognitive abilities (e.g. Tomasello, 1999). Pinker writes that if this were true, then *"these more general cognitive capacities would be the adaptation"* (p. 21, also see his quote in section 2.2.2). He nevertheless dismisses the possibility because the notion of 'general cognitive mechanisms' is too vague and thus hard to evaluate. Moreover, he claims that general learning mechanisms could not explain language universals, the uniformity and speed of child language acquisition, the emergence of creoles, and various kinds of language disorders.

The most interesting aspect of Pinker (2003) is that he proposes a number of tests to verify whether adaptation is a plausible process for the origins of language. One of them involves computational and mathematical models which can verify whether the proposed mechanism could indeed evolve some feature of language and under which circumstances this is possible. I will discuss these models in more detail in section 3.1.1. A second test involves molecular evolution, which should check whether selection really occurred. Pinker's theory basically makes two predictions with this respect: (1) there should be genes that are special to language, and (2) there should be many genes for language. The results in this field are far from conclusive, but advances in molecular biology may once indeed prove to be a fruitful way of falsifying or confirming the adaptation theory (for a review of the tests, see Pinker, 2003, p. 33–37).

2.5.2 Language as a lucky monster

Next to adaptation, ‘exaptation’ has also been suggested as a possible explanation for the origins of language. Exaptation is sometimes used as a wastebasket term for every evolutionary mechanism other than adaptation, but in its more narrow sense it means that something that has evolved for a particular function suddenly shifts to serve another function. One example would be the human respiratory system, which evolved for breathing but now also serves the function of articulating speech sounds. The idea that language evolved as such a ‘lucky monster’ is supported by some famous scientists in the field of evolutionary theory such as the late Stephen Jay Gould. For example Gould & Vrba (1982) argued that language essentially evolved as a by-product of increased brain size and other evolutionary forces.

Chomsky himself used to be vary vague about his view on the origins of Universal Grammar, but got placed in the exaptationist camp by Pinker & Bloom (1990) based on some of his quotes. Newmeyer (1994), however, claims that Chomsky does not exclude the adaptationist stance. Chomsky finally made an end to this speculation in a recent interdisciplinary study (Hauser *et al.*, 2002) in which – surprisingly – the hypothesis is put forward that only recursion is specific to language as part of the ‘narrow language faculty’, whereas other mechanisms are part of the ‘broad language faculty’ and are (at least partially) shared with other species. These findings are supported by comparative studies with animal communication systems. The article makes no claims about which evolutionary forces were crucial in the development of the language faculty, but raises the possibility that it may have evolved for other purposes than language and thus supports the exaptationist view.

The article caused a bit of a stir, not in the least because it is a radical break with Chomsky’s earlier claims about a complex and specialised language system. Some cognitive linguists read the paper as an example of a growing number of generativists admitting that many aspects of language can be attributed to general cognitive abilities (Goldberg, 2003). The article also led to a response from Pinker & Jackendoff (2005) who keep defending the adaptationist view. They argue that Hauser *et al.* (2002) ignore many features of language which are not recursive (e.g. case, agreement, etc.), and that the recursion-only claim is inspired by Chomsky’s Minimalist Program (Chomsky, 1995), which itself only features two basic operations ‘move’ and ‘merge’. Pinker & Jackendoff (2005) argue that the Minimalist Program ignores the same complexities of syntactic structures and is therefore inferior to earlier versions of generative grammar. The debate then continued in Fitch *et al.* (2005) and Jackendoff & Pinker (2005).

2.5.3 The recruitment theory of language origins

There is also a third explanation called the recruitment theory of language origins (Steels, 2007). The recruitment theory – which is actually more a research program than a fully-fledged theory – turns the research question of the origins of language upside down: instead of asking what genetic processes could have led to the present-day language capacity, it tries to identify what the minimal requirements are for the formation or emergence of certain linguistic phenomena. Steels hypothesises that *"the human language faculty is a dynamic configuration of brain mechanisms which grows and adapts, like an organism, recruiting available cognitive/neural resources for [...] achieving the task of communication"* (ibid., at 130). The recruitment theory thus tries to put the burden of grammar formation as least as possible on genetic evolution and rather focuses on *cultural* evolution and general cognitive mechanisms.

The recruitment theory can thus be placed in the family of theories that argue that grammar is developed through general cognitive mechanisms, which are criticised by Pinker (2003) for being too vague and too hard to assess. However, Steels (2007) argues that the same precise tests can be devised for the recruitment theory as for the hypothesis that language evolved as an adaptation: typological case studies, evidence from brain activity and biology, mathematical models, and finally computational models. In fact, there is already a decade worth of computational models testing the recruitment theory of language origins. Steels briefly gives examples of experiments that yielded successful results in the domains of lexicon formation, predicate-argument structure, and marking of egocentric perspective.

Alternatively to Pinker's prediction that language as an adaptation would involve many genes, the recruitment theory predicts that there are many mechanisms involved in the emergence of grammar which are not specific to language. One such example is egocentric perspective reversal (i.e. the ability to 'see' the world from a different perspective), which is used for a variety of non-linguistic tasks such as navigation. Two examples of perspective reversal in language are:

(2.9) The ball is to *your* right.

(2.10)	*kono+INFL*	*sono+INFL*	*ano+INFL*
	'near speaker'	'near hearer'	'away from speaker and hearer'

(Japanese – Kuno, 1973, p. 27, cited from Diessel, 2005)

Example 2.9 shows how the speaker uses lexical items to mark the visual perspective in her sentence. By saying *to your right*, the hearer knows that she can use her own perspective. Similarly, English speakers know that the demonstratives *here* versus *there* take the speaker as the point of reference. Example 2.10 shows how Japanese has grammaticalized a three-way contrast in its demonstratives, which alternatively take the speaker, the hearer, or both as the point of reference. For a small overview of spatial categories coupled to experiments with robotic agents, see Loetzsch *et al.* (2008).

Note that the recruitment theory does not exclude the possibility of language-specific genetic categories, but that it takes this explanation only as the last resort if all other pathways of investigation lead to a dead end. This dissertation subscribes itself to the

Generativist grammar	**Language as a complex adaptive system**
Language studied as an abstract system	Language serves complex communicative interactions
Innate grammatical categories	Language-specific, culturally constructed categories
- Static adult grammar - Child-based language evolution	- Grammar as an organically changing system - Emergent grammar
- Abstract rules versus lexicon - Autonomous syntax	- Continuum of constructions - Form-meaning pairs

Table 2.1: This table summarises the ongoing shift of attention in linguistics.

recruitment theory of language origins and tries to identify the minimal requirements for the formation of a case-like grammatical system to mark argument structure.

2.6 Summary: a novel conception of grammar

This chapter started with the most basic question that every linguistic theory should address: *why* do we use language? Knowing the function of language is the first step towards finding out why and *how* language originated in the first place. Based on the major impact of communicative pressures on grammatical constructions (as I will try to demonstrate more thoroughly in the rest of this thesis), I argued that language first and for all serves to establish social and communicative interactions between language users. From this view it follows naturally that every theory of language and grammar should focus on the complete communicative cycle, including (1) the participants of the communicative interaction (including their cognitive capabilities), (2) the communicative goal(s) of the dialogue (i.e. the functional pressures that have to be overcome), and (3) the setting in which the dialogue takes place (i.e. the real-world environment as well as the cultural setting).

The next question was how much of a speaker's linguistic knowledge is genetically given. The more the communication system is supported by innate structures, the less influence on the development of grammar can be attributed to external factors. I referred to an increasing body of empirical studies that suggest that the linguistic input observed by children is rich enough for them to construct a language by using general cognitive learning skills. These investigations are supported by cross-linguistic empirical data, which makes the hypothesis of universal grammatical categories extremely hard to defend. I therefore argued that grammatical categories are language-specific and that they are constructed culturally.

If grammatical categories are not prewired in the brain, then genetic evolution cannot explain where they come from. As an alternative, I proposed that grammatical categories emerged through the same mechanisms that are responsible for the present-day emergence of new grammatical forms (i.e. grammaticalization) and language change. I rejected the child-based view of language change because it implicitly assumes that grammars remain static during adulthood, whereas all empirical data suggests otherwise. I argued that grammar is continuously shaped and reshaped through language use, and that the best way to capture this observation is in terms of a cultural selectionist system.

Finally, I addressed the question as to which mechanisms are responsible for the origins and evolution of grammar. Next to referring to proposals made in the field, I subscribed the work presented in this thesis to the recruitment theory of language origins, which tries to identify the minimally required cognitive mechanisms for the development of grammar. I also suggested that the emergence of grammar cannot be attributed to a handful of cognitive operators, but rather that the language faculty may ‘recruit’ mechanisms from any other area of the cognitive-neural system depending on which problem needs to be solved.

None of the above proposals are completely new and credit is given to those scholars who pioneered them. Indeed, the work presented in this thesis is strongly aligned with cognitive-functional approaches to language in general, and construction grammar in particular. When taking all these aspects together, one may notice that there is an ongoing shift of attention in linguistics, abandoning the generative approach, which dominated the field for more than half a century. As is always the case, some of the assumptions of the old tradition keep lingering around unnoticed for a long time (e.g. the focus on acquisition as the sole driving force of language change or cross-linguistic studies which look for categories which are obsolete from a functional point-of-view), so it is (still) necessary to make your definition of grammar explicit.

Chapter 3

Modeling the formation of language

The experiments reported in this thesis use a methodology which is related to research in the field of Artificial Life (or Alife). Rather than trying to simulate actual natural phenomena found in real life, Alife researchers use computational simulations, robotic experiments and biochemistry to create (new) artificial phenomena that have similar characteristics as some real-life phenomena and hence provide possible and operational explanations for them. The growing number of Alife articles in top journals such as *Nature* and *Science* shows that the field is becoming more and more accepted as implementing a valid research methodology. Techniques of artificial language evolution, however, are still relatively unknown in (computational) linguistics. This is mainly due to the fact that attempts at coupling experimental results to linguistic theory are few and far between. One of the goals of this thesis is therefore to take a first step at mending this scanty engagement with theory.

In this chapter, I will first give a brief overview of the three main approaches to modeling the evolution of grammar found in the field: genetic evolution, cross-generational cultural evolution, and negotiation-based cultural evolution. I will argue that the latter account, also known as the 'complex adaptive systems' or 'problem-solving' approach, is the most fruitful way to go about it. Next, I will briefly discuss the methodological standards that experiments on artificial language evolution have to live by if they are conducted within a cognitive-functional framework. Finally, I will give a short history of the results obtained so far in this line of work.

3.1 Three approaches to modeling grammar

Languages change. Speakers may not always be aware of this fact – changes are usually small and they take a lot of time to become conventionalised in a language – but changes nevertheless occur. This becomes apparent when students are confronted with literary works of a distant past. Take for example the following line:

(3.1) *Nu sculon heriġean \\ heofonriċes Weard*
(Abrams & Greenblatt, 2000, p. 24)

Speakers of present-day English may find these words as alien and exotic as they might appreciate Swahili poetry, but they actually come from *Cædmon's Hymn*, probably the oldest extant poem in the history of English literature (written somewhere between 658 and 680). It translates as 'Now we must praise the Guardian of the kingdom of heaven (i.e. God)' and some words can be understood with a bit of imagination: *nu* 'now', *sculon* 'shall/must', *heofon* 'heaven', and *weard* 'ward/guardian'. Speakers of Dutch might also recognise *-ric* '-rijk' as in *hemelrijk* 'kingdom of heaven'. As for the grammar, the subject is not explicitly mentioned, but the modal auxiliary *sculon* is marked for agreement through the suffix *-on*. The word *heofonric* is marked by the genitive case marker *-es*, which is still remnant in today***'s*** English.

Research on the formation of language has always paid special attention to the processes of language change and there is a wide consensus that language must have evolved and still continues to evolve because there is a selectionist system underlying it. Despite this consensus, there is disagreement on how variation and hence the potential for change is caused and what selectionist pressures are operating to retain a particular variation in the language. There are basically three different approaches based on genetic evolution, cultural transmission, and problem-solving respectively. All three also have different opinions of what the primary trigger is of grammaticalization. In the remainder of this section I will illustrate the difference between the three approaches through the debate of 'holistic' languages (which have a different word for each complex meaning) versus 'compositional' languages (which reuse smaller units to compose novel utterances).

3.1.1 Models of genetic evolution

The genetic evolution models (e.g. Briscoe, 2000; Nowak *et al.*, 2001; Nowak & Krakauer, 1999; Nowak *et al.*, 1999) put the selectionist pressure at the level of fitness, which is assumed to be directly related to communicative success. Agents are endowed with an artificial genome that determines how they should communicate: what perceptual primitives they should use for segmenting the world and identifying objects and features, what concepts they can employ for structuring their world, what types of categories are to be used, etc. Potential innovation takes place at the moment this genome is transmitted from parents to children. Because genome copying involves crossover and possibly mutation, variation is inevitable, and some of it will lead to higher or lower success. Models of genetic evolution have the most affinity for linguistic theories supporting 'Universal Grammar' (ranging from strong nativism such as the P&P model to the generalised notion of UG, see section 2.2.1) or the 'language as an adaptation' hypothesis (Pinker, 2003; Pinker & Bloom, 1990; Pinker & Jackendoff, 2005, see section 2.5.1). I will briefly discuss the model of Nowak & Krakauer (1999) to illustrate this approach.

Selectionist system. Nowak & Krakauer (1999) report a mathematical model in which agents are engaged in a series of 'evolutionary language games' (Nowak *et al.*, 1999). In each round of the game, every agent speaks to every other individual in the population and both the hearer and the speaker receive a reward if the game was a success. At the end of each round, the individual 'payoffs' are counted for each agent. Following evolutionary game theory, this payoff is interpreted as a fitness function: speakers with the highest payoff have a higher fitness score and thus increase their chances of producing offspring.

Next to models of the emergence of signal-object associations and word formation, Nowak & Krakauer (1999) also write about an evolutionary language game in which agents have to talk about simple events. These events consist of one-place predicates involving one action (A) and one object (O). A combination of an action and an object is called an event so the same action can combine with different objects to yield different events.

The speakers in the experiment can basically evolve two different strategies for expressing events: a 'non-grammatical' (i.e. holistic) one in which they introduce a new word for each meaning, or a 'grammatical' (i.e. compositional) one in which nouns and verbs can be combined by a syntactic rule. Here is an example of such a grammatical language with only two objects (O) and two actions (A), which can be expressed by nouns (N) and verbs (V):

(3.2)

O_1A_1	N_1V_1
O_2A_1	N_2V_1
O_1A_2	N_1V_2
O_2A_2	N_2V_2

The trigger of grammar. The fitness function or payoff is the crucial factor that in the end will favour either one of the above two strategies. Nowak & Krakauer introduce the hypothesis that the emergence of grammar is (mainly) driven by inventory optimisation. This hypothesis has an intuitive appeal to many linguists because it fits into the traditional idea that language users can produce an infinite set of sentences using a finite set of building blocks. The capacity of recursion is thus seen as the essential feature of grammatical languages.

Following this hypothesis, a compositional language receives a higher fitness function than a holistic one once the number of events (combinations of actions and objects) exceeds or equals the number of nouns and verbs that can be used to describe the events. In the above example, this is already the case: there are four events possible requiring two verbs and two nouns. The number of events is thus the same as the number of words used. Speakers that use the compositional strategy thus have a higher fitness score than those that use the holistic one, and therefore have a higher chance of producing offspring.

Discussion. Nowak & Krakauer (1999) conclude that *"the need for grammar arises only if communication about many different events is required: a language must have more relevant sentences than words"* (p. 8031). Even though syntax indeed has an optimisation effect, this conclusion is an overinterpretation of the results due to the model's poor semantics. Consider the following two examples:

(3.3) Push boy girl.

(3.4) The girl pushed the boy.

Example 3.3 illustrates the problem that you get once speakers start reusing words in different utterances rather than opting for the holistic strategy: it is unclear whether it was the boy who did the pushing in this event or whether it was the girl. The sentence might also be interpreted as the boy and the girl pushing something together, or that both were being pushed, or that the boy pushed something to the girl or vice versa, etc. Even though the sentence is optimal from the point of view of the inventory size, it has some serious disadvantages with respect to the holistic strategy: the hearer needs to figure out for herself how the meanings of the individual words are linked to each other and even needs to witness the scene in order to do this.

Nowak & Krakauer's model has the implicit (but fundamental) assumption that this linking of meanings is automatically taken care of by syntax (or grammar). But why, from the viewpoint of optimising the inventory, would a speaker ever want to do this? The answer seems to be that the speaker goes through a lot of pain to make the interpretation task easier for the hearer. Even in a simple sentence such as example 3.4, the speaker uses grammar to mark that:

1. the girl was the pusher in the event and also the topic of the utterance (this is achieved by placing 'the girl' in the subject position of an active clause and using subject-verb agreement);

2. the boy was the one being pushed in the event (marked by the object position in the active clause);

3. both the girl and the boy are uniquely identifiable objects in the speech setting (both marked by the determiner *the*);

4. the action took place in the definite past with respect to the moment of speaking (marked by the *-ed* participle on the verb, which here indicates the simple past tense).

The hypothesis that grammar is needed only to optimise the linguistic inventory can thus only be a drive for evolving lexical compositional languages rather than grammatical compositional languages unless the grammar is built-in completely. A lexical language may be an important prerequisite for developing a grammatical language, but it does not directly cause it since one can (often) infer the correct meaning from the context.

3.1.2 Iterated Learning Models

Iterated Learning Models (ILM, Brighton *et al.*, 2005a; Kirby, 2000, 2001; Kirby & Hurford, 2002; Kirby *et al.*, 2004; Smith *et al.*, 2003) are similar to genetic models in the sense that variation and hence potential innovation takes place in the transmission of the language system from one generation to the next, but now the language and conceptual system are considered to be culturally coded instead of genetic. Children learn the language from their parents and are then assumed to use it largely unchanged throughout the rest of their lives. The learning process necessarily introduces generalisations and variations because of the Poverty of the Stimulus, and hence innovations may enter into the acquired language system. This innovation may re-appear in the data that the learners generate for the next generation once they have become adults and thus gets preserved. The ILM features a model of language change which is almost identical as the one illustrated in Figure 2.1 in section 2.3.1; and assumes a Universal Grammar in its more generalised sense as defined in section 2.2.1.

Before I illustrate the model's approach to the debate on holistic versus compositional languages, it should be noted that some researchers have generalised the model to include intra-generational transmission as well (e.g. Smith, 2003a). In order to avoid terminological confusion, I will continue to use the notion of 'Iterated Learning Model' in its cross-generational transmission sense since it represents the bulk of its research program. I will classify those models that *do* take communicative pressures and single-generational dynamics into account as belonging to the third approach in which language is seen as a complex adaptive system (see section 3.1.3).

Selectionist system. The most common population model found in ILM experiments is that of one adult speaker (representing the adult generation) and one child learner (representing the younger generation). In the experiments, the adult speaker produces a series of utterances which are observed by the child. The learner is equipped with a Universal Grammar in the form of an inductive learning algorithm and tries to reconstruct the adult grammar based on the observed utterances. There is thus no direct link between the adult's grammar and the child's grammar. After a while, the adult speaker 'dies' and the child becomes the new adult speaker. A new agent is introduced and the whole cycle starts again and this process iterates over thousands of generations.

In experiments on the emergence of compositional and recursive syntax (Kirby, 2000, 2001; Kirby & Hurford, 2002; Brighton *et al.*, 2005a,b), the adult speaker has to describe a number of scenes to the child learner. If the speaker has no way yet to express a certain (complex) meaning, she will invent a random and holistic string for it. The child, however, may overgeneralise the input when the same character or sequence of characters accidentally occurs in two utterances which partially overlap in meaning. Variation in the model thus comes from mismatches between the adult and child grammar due to learning. Once the child becomes the adult speaker, the overgeneralisation may be passed on to the next generation instead of the original holistic form.

The trigger of grammar. Most of the experiments in the ILM follow what Brighton *et al.* (2005a) call the 'function independence principle', which states that not all structure should be assigned to communicative pressures but rather to iterated transmission. The trigger of grammar therefore has to be found in pressures on transmission. The most important pressure is hypothesised to be the 'Poverty of the Stimulus'.

This hypothesis is implemented in the experiments in the form of a 'transmission bottleneck'. The meaning space offered to the agents is significantly larger than the number of utterances they can observe during a lifetime. The holistic form for a particular meaning may therefore get lost during the cultural transmission of the language from one generation to the next simply because the child learner did not observe it. Words that cover only partial meanings, however, reoccur more frequently and therefore have an evolutionary advantage for being learned by the new agent. In other words, smaller linguistic items which combine into utterances are more frequent and are thus observed more by the learner. An example of such a compositional, recursive grammar is offered by Kirby & Hurford (2002, p. 131–132):

(3.5)

$$\begin{array}{c} S/p(x,y) \rightarrow \texttt{gj}\ A/y\ \texttt{f}\ A/x\ B/p \\ S/p(x,q) \rightarrow \texttt{i}\ A/x\ D/p\ S/q \\ A/\text{heather} \rightarrow \texttt{dl} \\ A/\text{mary} \rightarrow \texttt{tej} \\ A/\text{pete} \rightarrow \texttt{n} \\ A/\text{gavin} \rightarrow \texttt{gp} \\ A/\text{john} \rightarrow \texttt{h} \\ B/\text{detests} \rightarrow \texttt{b} \\ B/\text{loves} \rightarrow \texttt{wp} \\ B/\text{hates} \rightarrow \texttt{c} \\ B/\text{likes} \rightarrow \texttt{e} \\ B/\text{admires} \rightarrow \texttt{m} \\ D/\text{believes} \rightarrow \texttt{g} \\ D/\text{knows} \rightarrow \texttt{u} \\ D/\text{decides} \rightarrow \texttt{ipr} \\ D/\text{says} \rightarrow \texttt{p} \\ D/\text{thinks} \rightarrow \texttt{m} \end{array}$$

This syntax can lead to utterances such as the following (ibid., at 131, examples 12, 13 and 16, spaces are for clarity and not given to the agents):

(3.6) *gj h* *f tej* *m*
John Mary admires
'Mary admires John.'

(3.7) *gj h* *f tej* *wp*
John Mary loves
'Mary loves John.'

(3.8) *i h* *u* *i tej* *u* *gj qp* *f h* *m*
John knows Mary knows Gavin John admires
'John knows that Mary knows that John admires Gavin.'

The Poverty of the Stimulus as a trigger in itself is not sufficient however. The Iterated Learning Model can only function properly (a) if there is a learning bias or Universal Grammar so that the agents will overgeneralise the input, and (b) if the meaning space itself is rich and (recursively) structured.

Discussion. So what about functional pressures? Brighton *et al.* (2005b, p. 197–109) write that *"there is no need to appeal to this notion of functionality [i.e. communicative pressures] to explain the cultural evolution of compositional structure – compositionality evolves in response to the pressure introduced by the transmission bottleneck, and it is therefore unnecessary to invoke an explanation which appeals to the combination of a transmission bottleneck plus a pressure for communication."*

Brighton *et al.* (2005b) then argue that the above conclusion is correct only if the emergent compositional languages are also functional for communication. They therefore measure communicative accuracy as well, which is here a global measure: accuracy is defined as the probability, averaged over all meanings, that the speaker will produce an utterance for a certain meaning which will be interpreted correctly by the hearer. The results show that the languages indeed evolve into a syntactic structure which leads to optimal communicative accuracy.

This notion of communicative accuracy is however highly problematic: since it is a global measure, communicative accuracy does not bear the slightest impact whatsoever on the linguistic behaviour of the agents (as was indeed intended according to the function independence principle). This allows for a serious mismatch between the internal states of a speaker and a hearer: in the results reported by Brighton *et al.* (2005b, p. 198, fig. 5), communicative accuracy is only reached after twenty generations. So even though the agents engage in communicative interactions, they don't care about communicative success. It is therefore highly questionable whether the transmission bottleneck still plays such a major role if the agents would have to negotiate the meaning and try to reach mutual agreement already in their local interactions (also see the next section).

For the reasons given in section 2.3.2, the ILM also makes the wrong predictions about language change when we compare the results to the evolution of natural languages. This criticism seems to be taken seriously as there is a recent shift from using the ILM as a model for present-day language to a model for 'proto-language', a simpler precursor of modern languages (e.g. Jackendoff, 2002). While reviewing evidence that the invention of new words and grammatical forms does not seem to involve a holistic precursor, Smith (2008) suggests that the most dominant mechanism of change may have shifted from 'analysis' (as proposed by the ILM) to grammaticalization.

Demonstrating when, why, and how such a shift may have come about in language change poses a serious problem on the model and goes against the hypothesis of continuity in grammaticalization (see Bybee's quote in section 2.3). But it is not the only problem left: just like the genetic model of Nowak & Krakauer (1999), some ILMs can only predict the emergence of a lexical language (e.g. Smith *et al.*, 2003). The meaning space seems to be more complex than Nowak & Krakauer's model, but the models can in fact not distinguish between *Jack kisses Mary* and *Mary kisses Jack* unless a different form is used for each word depending on whether it was used as an agent or as a patient (see De Beule & Bergen, 2006, for further discussion). One exception is Kirby (2000) in which word order is used. But here semantic roles such as 'agent' and 'patient' are already built in and agents expect word order to be meaningful. Moreover, recent attempts by Moy (2006) to get rid of word order in favour of a grammar for case have failed. Since the research question addressed by Moy (2006) is roughly the same as the one tackled in this dissertation, I will come back to this matter in Chapter 8 where I offer a more thorough comparison.

A final reason for rejecting the learning bottleneck as the primary trigger for grammar is that the Iterated Learning Models only work if the agents are equipped with a strong inductive generalisation mechanism. The learner *expects* its input to be structured so the preference for compositionality is already built in. This in fact suggests that compositionality is due to genetic evolution rather than cultural transmission.

3.1.3 Problem-solving models

The third class of models views the task of building and negotiating a communication system as a kind of problem-solving process. Agents try to achieve a communicative goal with maximal success and minimal effort. This problem-solving process is definitely not a rational but an intuitive one that is seldom accessible to conscious inspection. It is not an individualistic problem-solving process either, but a collective one, in which different individuals participate as peers. According to this view a communication system is built up in a step by step fashion driven by needs and failures in communication, and it employs a large battery of strategies and cognitive mechanisms which are not specific to language but appear in many other kinds of cognitive tasks, such as tool design or navigation. Recent experiments by Galantucci (2005) on the emergence of communication in human subjects provide good illustrations of these problem-solving processes in action. Variation and innovation in problem-solving models are common because each individual not only tries to converge to the shared communication system, but can also contribute innovations to it. In fact the main challenge is rather to explain how agreement between individuals and thus a globally shared population language can ever arise. This approach to language evolution is closely related to cognitive-functional theories of language and utterance-based models of linguistic change (Croft, 2000, 2004).

Selectionist system. In the problem-solving approach, language becomes a Complex Adaptive System (CAS) in its own right, similar to a complex ecosystem or a complex economy (Steels, 2000b). There are many parallel competitions going on: between synonyms for becoming dominant in expressing a particular meaning, between idiomatic patterns that group a number of words, between different syntactic and semantic categories competing for a role in the grammar, between ways in which a syntactic category is marked, etc. An innovation only survives if it is successful in communication (which could be due to many factors such as the effectiveness of the meanings involved) and if it is also picked up and further propagated by a sufficient number of agents. Often there is no particular reason why one solution is preferred over another one, except that it is more frequent in the population and it wins because of the rich-get-richer dynamics. So we get two types of selectionist forces: functional and frequency-based, which are also the key forces of language change proposed by usage-based theories of language (Croft & Cruse, 2004, chapter 11). Also note that, as opposed to genetic and Iterated Learning Models in which there is no agent-internal competition, every agent may know several variations with various degrees of conventionalisation at any given time (also see section 2.3.2).

This approach has also been applied to the debate on holistic versus compositional languages by De Beule (2007, 2008); De Beule & Bergen (2006); and to a lesser extent by Wellens *et al.* (2008). I will briefly illustrate the approach by summarising the work of De Beule & Bergen (2006). This experiment features a single-generation, multi-agent population in which each agent engages both as a speaker and as a hearer in peer-to-peer interactions (called 'language games'). There is thus no genetic transmission, nor a cross-generational turnover. Each language game, two agents are randomly chosen to act as either the speaker or the hearer. The speaker uses her linguistic inventory (if possible) to describe an event to the hearer, or will invent new forms if necessary. Variation arises naturally in the model because the agents start without any language and different agents may therefore invent a different form for the same meaning.

The agents store each variety they invent or learn in their memory and assign a 'success score' to it which reflects how successful this particular form-meaning mapping has been in past communicative interactions. If it was successfully used, its score increases and the scores of the competitors are decreased through lateral inhibition. If the language game was a failure, its score is decreased. Smaller linguistic items in De Beule & Bergen (2006) get introduced as innovations by the speaker for less complex meanings rather than being induced by a hearer's biased learning mechanism. Holistic and compositional encodings now start to compete with each other within the population and even within one single agent. In the end, compositional structures win the competition: once smaller items are acquired they can be used in more situations and are thus more frequent and will be reinforced more often. The holistic words have to be learned each time and have only one use so they are rarely reinforced. The agents thus have no particular preference built in for either holistic or compositional forms but will opt for the ones with a higher estimated communicative success.

The trigger of grammar. There is a big difference in meaning when we say *Jack pushed Jill* versus *Jill pushed Jack*. In the genetic and Iterated Learning Models, this difference in meanings was either ignored (i.e. the meaning space was too simple) or handled by a built-in Universal Grammar (i.e. meanings are assumed to be linked to each other through word order). By ignoring semantic complexity and communicative pressures, the other models overinterpreted inventory optimisation and transmission bottlenecks as the pressures for grammaticalization. The work on language as a complex adaptive system takes a different stance and looks for the triggers of grammar in function and communication.

When language users engage in a communicative interaction, the speaker wants to achieve a certain effect in the hearer. There are many intended effects possible depending on the speech act, such as drawing the hearer's attention to a particular object or event. In order to achieve the intended effect and thus reach communicative success, the speaker will try to reduce the complexity of semantic interpretation for the hearer and try to avoid ambiguity (Steels, 2000a, 2005b; Steels & Wellens, 2006). Applied to the experiments reported in De Beule & Bergen (2006), the agents will not automatically induce word order rules but only introduce new constraints when it is useful for communication. This happens when a speaker wants to reuse existing words into a new compositional utterance: there may be ambiguity as to who did what to whom in the event so the speaker might introduce additional grammar to solve this problem. The hearer too will use her previous knowlegde, cognitive repair strategies and the situatedness to infer what the speaker meant.

Discussion. De Beule & Bergen (2006) show that compositional languages can emerge if communicative success and function are taken to be the driving forces behind language evolution. Their model improves over genetic and Iterated Learning Models because no explicit preference for compositionality is built in (only the indirect bias to reuse existing items to reach success in the game) and only one generation is required to reach a stable and shared communication system. The model also acknowledges that grammar has to link meanings to each other and offers a functional trigger for the development of more grammatical structures.

The three approaches discussed here are of course not mutually exclusive: it is clear that the models of cultural evolution expect a rich cognitive-linguistic system (e.g. the capacity to build composite structures), which can only be explained through the genetic endowment of the agents. Similarly, problem-solving models can be modeled using a generation turnover as well (Vogt, 2007). However, focusing exclusively on genetic evolution or on transmission effects leads to an overinterpretation of experimental results as I argued earlier in this section. The main advantage of the problem-solving approach is that it attempts to explain as much as possible in terms of functional and communicative pressures and only uses innate mechanisms as a last resort (which arguably leads to a more plausible hypothesis of what is innate and what is not). I will elaborate further on this claim in the next section.

3.2 The do's and don'ts of artificial language evolution

In 1866, the *Société de Linguistique* in Paris banned all research on the origins of language. The ban was installed because papers on this topic were too speculative and based on little or no empirical data: there are no fossil traces or written accounts left of the first languages, and there is no data available about the biological changes that enabled language. These facts have led Christiansen & Kirby (2003) to pose the provocative question: is language evolution the hardest problem in science?

Without attempting to answer that question, it actually *is* possible to come up with solid scientific theories despite the lack of real-life data: in various other disciplines, such as the origins of the universe, significant advances are made using mathematical models and computational simulations. Another example is research in Artificial Life where computational and analytic models, robotic experiments and biochemistry are used as techniques to develop (new) artificial phenomena which may help to explain processes of evolution in natural phenomena. These techniques are however novel to (computational) linguistics, so a brief explanation of their methodology is warranted.

In this section I will discuss how to do experiments on artificial language evolution and why (for a more thorough discussion, see Steels, 2006). I will also point to the limits of the approach and how it should be coupled to evidence gathered in other disciplines. This discussion is however only valid for studying the formation of language in a problem-solving approach, which I will adopt in this thesis (see section 3.1.3). The other approaches (whether genetic or Iterated Learning Models) may have different standards and measures for assessing progress made in the experiments and I do not suggest that these models should follow all the criteria proposed here in order to yield valid and relevant results. Most of the issues discussed here are however shared by all approaches and also by other scientific disciplines.

3.2.1 Creating natural language-like phenomena

Let me start by debunking a myth about modeling language evolution that seems to be lingering in the minds of some people: computational simulations and robotic experiments can never lead to the emergence of an actual natural language like English or Russian. There are just too many factors that have shaped these languages and modeling them would require to model the entire state of every speaker and every interaction *ever*. The methodology is thus neither appropriate for making predictions of actual change in natural languages, nor for tracing back the history of some existing language. The 'languages' that emerge in this kind of experiments should thus not be *directly* linked to natural languages: they are novel artificial constructs and hence **the methodology only offers a 'proof of concept' of what results can be achieved within the specific set-up of a given experiment**. However, these constructs may be natural language-*like* and thus offer a possible and working hypothesis for how similar phenomena could have come about in natural languages.

Another limit of the methodology is that even in abstraction, not every aspect of language can be modeled at once, just as it is impossible to perform controlled psycholinguistic experiments that take every detail about language processing into account. The key is therefore to pick a phenomenon observed in natural languages and try to isolate this feature into a controlled simulation or experiment. Focusing on smaller problems is standard practice in many scientific disciplines. As Stephen Hawkin writes in his *A Brief History of Time*:

> It turns out to be very difficult to devise a theory of the universe all in one go. Instead, we break the problem up into bits and invent a number of partial theories. Each of these partial theories describes and predicts a certain limited class of observations, neglecting the effects of other quantities, or representing them by simple sets of numbers. It may be that this approach is completely wrong. If everything in the universe depends on everything else in a fundamental way, it might be impossible to get close to a full solution by investigating parts of the problem in isolation. Nevertheless, it is certainly the way we have made progress in the past.
> (Hawkin, 1988, p. 12)

In cognitive-functional models of artificial language evolution, the researcher therefore first chooses a topic of interest that is common to most if not all languages, such as tense, aspect and mood. Investigating such a feature involves the following steps (Steels, 2006):

1. The researcher selects a feature of language to investigate;
2. The researcher hypothesises which set of cognitive mechanisms and external pressures are necessary for the emergence of this feature:
   (a) These cognitive mechanisms are operationalised in the form of computational processes, and a population of simulated or embodied agents is endowed with these mechanisms;
   (b) The external pressures are operationalised in the form of an interaction pattern embedded in a simulated or real world environment;
3. Systematic computer simulations are performed demonstrating the impact of the proposed mechanisms. If possible, results should be compared between simulations *with* a proposed mechanism and simulations *without* this mechanism in order to show that it is not only a sufficient but also a necessary prerequisite for the emergent feature.

Even if the simulations show that the investigated feature only emerges if certain mechanisms are included, this still does not prove anything about similar features in natural languages because different evolutionary pathways are possible. However, the simulations then at least show one possibility and may provide an additional piece of the puzzle next to evidence from linguistics, archeology, biology and other fields.

3.2.2 Clarifying the scaffolds and assumptions

Computational simulations and robotic experiments are at the same time blessed and cursed with the fact that every detail of the hypothesis has to be spelled out completely, otherwise the system does not work. 'Blessed' because this may reveal effects of the hypothesis that were overlooked or not expected by the verbal theory; 'cursed' because there is a danger of 'kludging' something together to get the system off the ground. There is often a fine border between significant results and quirks resulting from a kludge, so it is crucial that it is crystal clear which features of the model are supposed to be 'emergent' on the one hand, and which features are 'assumed' or 'scaffolded' on the other.

Mixing up the features that are emergent with those that are given leads to wrong conclusions and overinterpretation of experimental results. For example, the Iterated Learning Models discussed in section 3.1.2 claimed that 'compositionality' emerges as the result of a transmission bottleneck from one generation to the next. However, child learners were equipped with a grammar learning device that gave them no choice but to induce compositional structures from (possibly) holistic input. Compositionality was thus at least partially given by the experimenter.

Assumptions and scaffolds are unavoidable in computational simulations because it is simply impossible to explain everything at once. Drawing the line between what is 'assumed' and what is 'scaffolded' is not always an easy decision. I define 'assumed features' as those aspects of the simulations that we cannot explain using the methodology described here. One example is the assumption that agents are social and cooperative beings that *want* to reach communicative success. Another example is the capacity for building composite structures. I define 'scaffolded features' as those aspects of the simulations that in principle could be evolved using the methodology but are given so that not every simulation or experiment has to start from scratch. An example of a scaffold is the space of possible phonemes: in all of the above experiments on compositionality, none of the agents had to worry about which sounds were distinctive in their language. These 'scaffolded' features may be brought in at a later stage or form the subject of other experiments, such as work on the emergence of vowel systems (de Boer, 2000; Oudeyer, 2005).

3.2.3 Global versus local measures

Another important dichotomy is that of **global versus local measures**. In experiments that study language from a usage-based point-of-view, only local measures which can be observed by the agents themselves within the local interaction may have an influence on the linguistic behaviour of the agents. Typical local measures are:

- Success of the language game: the agents that are involved in a language game can experience whether the game was a success or not. This may influence the confidence with which they employ certain linguistic items;

- Search and difficulty: An agent can 'measure' for example the ambiguity of an utterance because it causes an elaborate search space during processing;
- Cognitive effort: Agents can 'measure' the cognitive effort needed during parsing and production, such as how much processing time they need, how many times they need to add information from their world model to the linguistic data, how many times they need to perform additional operations such as egocentric perspective reversal, etc.

Global measures, then, are measures which are used by the experimenter solely for analysing the simulations. These measures should by no means have an influence on the linguistic behaviour of the agents. For example, if an agent should decide between two competing forms for a meaning, she has to make the appropriate choice based on her *individual* past experiences and on the local information in the language game. A global measure such as how many agents share one of the competitors would require an overview of the complete population, which no speaker ever has. Convergence has to come about in a distributed, self-organised fashion without external guidance or central control.

Global measures which guide experiments like a 'hidden hand' are most frequently found in genetic models of the origins of language. For example, Nowak & Krakauer (1999) report a model on word formation in which the fitness function is defined as the total amount of successful information transfer. The agents are however assumed to know the lexicon of the population perfectly and a correct understanding of a word depends on comparing the perceived word with all the words of a lexicon. So the model does not only use global measures (nobody ever knows the total number of words shared in a population), it also mixes up given versus emergent properties: the claim is that word formation emerges through natural selection, but the words are already given so only the sound-form matchings have to be acquired.

Assumptions like the ones proposed by Nowak & Krakauer are downright unrealistic, which may be one of the reasons why computational simulations and mathematical models are not always taken seriously in linguistics (and other disciplines). The field of the origins and evolution of language is fairly young, so effort has to be put into tightening the assessment criteria for sound application of the methodology. A first step is to communicate all the measures clearly, which I will try to do in this thesis.

3.2.4 The design stance

It should be clear by now that the experiments reported in this thesis are conducted from a **design stance**. As is the case for all studies using language technology, there are two ways of interpreting the notion of 'design stance' and the experimenter has to make clear which one is followed. The first way is the engineer's approach in which the experimenter is chiefly interested in a practical application to solve a certain problem (such as building an open-ended communication system for robots). The engineer will aim for the best performance and does not care about the relevance of the system for describing natural languages. The cognitive scientist's approach, on the other hand, *does* care about the coupling between the design and the natural phenomenon, and a comparison between the two may lead to a change in the design. In this dissertation, I commit myself to the latter approach and try to obtain relevant results for linguistics. Since the design stance makes different kinds of abstractions and has different objectives than verbal theories and other approaches in computational linguistics, the relevant insights of the models are very different in nature. I will go into more details about what this means for computational linguistics in Chapter 5 and for linguistics in general in Chapter 8.

As argued earlier, the design stance is very well suited for operationalising cognitive mechanisms and external factors and demonstrating their impact (in isolation from other forces) on the whole communication system. This also allows for identifying the **transition phases** of language development and showing how the transition from one phase to the next can be achieved. For example, if agents need to communicate about certain objects in their environment, a good strategy might be to give individual names to each object. However, this solution might become problematic if there are too many objects or if objects cannot always be uniquely identified. The agents could then start constructing more general categories or words that are not unique to each object, but can help to discriminate them in a certain context. The design stance allows the researcher to demonstrate in which cases a certain strategy is sufficient for reaching communicative success, and then to demonstrate the critical moment in which the chosen strategy becomes insufficient. The transition from the present stage to the next one can then be achieved by recruiting additional cognitive mechanisms and strategies. Some possible transition phases are illustrated in Figure 3.1.

The operationalisation of all the hypotheses and claims made in the design stance also goes through various stages of complexity. The first time, everything is tested in computational simulations that feature a predefined virtual world and a possible meaning and form space. Computational simulations are much easier to implement than robotic experiments so they offer the experimenter a faster way of testing her hypotheses for the first time. In a next step, the robustness of these simulations should be tested by implementing them into physically embodied agents (e.g. robots). Trying out the same experiments in real-world conditions is no trivial matter and involves more than simply connecting some vision system to the model. Finally, computational simulations and robotic experiments have the major drawback that they cannot make any claims about the long-term behaviour of the system. They therefore have to be complemented with analytic models that can measure mathematically what the impact of each parameter is on the system dynamics (Steels, 2006).

In order to be relevant to other research fields, researchers who follow the design stance have to engage in an interdisciplinary dialogue because there are too many questions which cannot be answered using this methodology: how did the human language

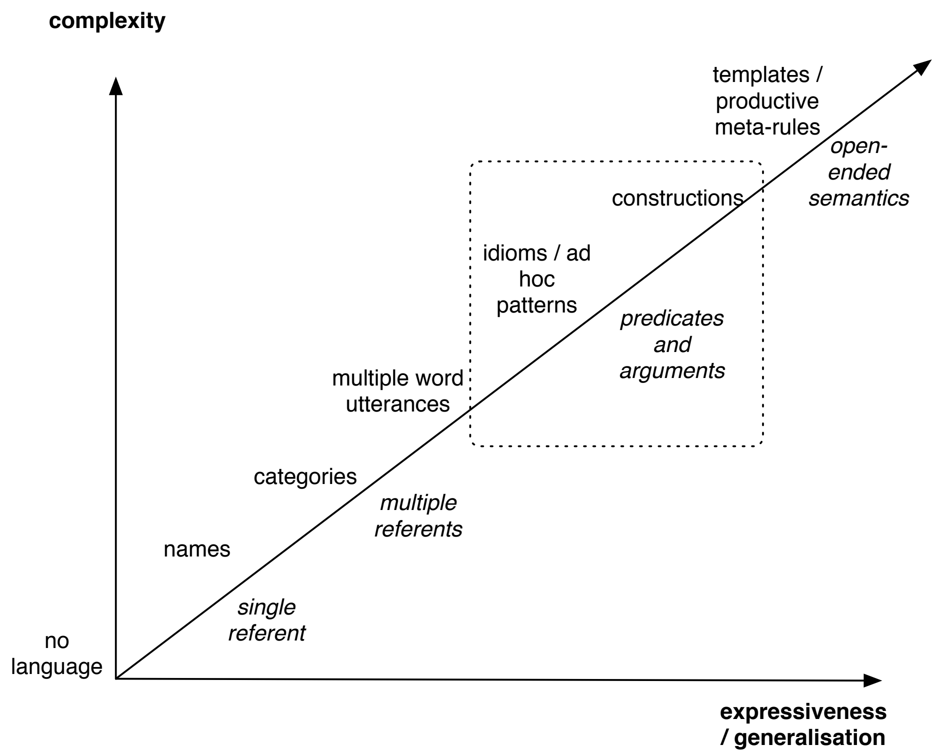


Figure 3.1: The design stance allows researchers to identify the transitions in language towards increasing complexity and expressiveness. This thesis focuses on the phase of idiomatic constructions and constructions to express predicate-argument structures.

capacity evolve? How is language neurobiologically processed? When did the first languages emerge? Is there evidence for the existence of the proposed mechanisms in humans? In this thesis I will attempt to couple the experiments to linguistic theory and to make the results accessible for linguists interested in computational modeling.

3.3 A brief history of the research

This dissertation is part of more than a decade worth of research on language as a complex adaptive system. The research itself is rooted in prior work in Artificial Intelligence and robotics in the mid-eighties, which involved home-made robots of all shapes and sizes – driving around on wheels, flying with balloons and propellers, or waggling their tails in the rough waters of the Brussels' university swimming pool. These creations were the result of a break with mainstream research in Artificial Intelligence: whereas most work in AI tried to formalise human intelligence, Luc Steels and his students at the Artificial Intelligence Laboratory (VUB, Brussels) investigated how 'intelligence' might evolve in a community of physical agents as they autonomously interact with each other, their environment, and humans. They thus developed a bottom-up, behaviour-oriented approach to sensori-motor intelligence, which was at the same time also being explored by Rodney Brooks at the MIT Artificial Intelligence Lab (Steels & Brooks, 1995).

Even though fascinating results were obtained, something was missing that could ever lead the experiments to other, more human-like intelligent behaviour than was displayed by the robots. The research then shifted its focus from behaviour-based robotics to embodied language in the autumn of 1995 when Steels had the following two ideas: first of all, language may have been the missing link in the initial experiments. Language may be a necessary step that enables the human cognitive system to bootstrap itself in tight interaction with the world and in a population of social cooperative agents (Steels, 2003a). Second, the same principles and mechanisms that had since 1985 proven to be relevant for the work in robotics also had to be relevant for bootstrapping intelligence and language. These principles included self-organisation, structural coupling, level formation and other (mainly biologically inspired) mechanisms (Steels, 1998b). In this section I will give an overview of the research efforts at the AI Lab in Brussels and SONY Computer Science Laboratory Paris (which adopted the work on language as one of its founding research topics in 1996).

3.3.1 The formation of adaptive lexicons

The first breakthrough experiment investigated how **self-organisation** could explain the formation of coordinated vocabularies (Steels, 1997e). Self-organisation is a phenomenon in which coupled dynamical systems form a structure of increased complexity without guidance by an outside source or some central controller. Standard examples of self-organisation are the formations of termite nests or paths in an ant society. The process has been used successfully for describing phenomena in various scientific disciplines: physics (e.g. crystalisation), chemistry (e.g. molecular self-assembly), economy (catallaxy), etc. Language is also an example of a complex system which is shared by a speech community without central control (even though some 'watchdogs' such as the *Académie française* do their utter best to have people speak and write according to their set of standards).

In the experiments reported by Steels (1997e), agents engage in communicative interactions about a set of predefined meanings. If the speaker has no word for a meaning, she will invent a new one which may be adopted by the hearer. The agents assign a **success score** to the form-meaning mappings based on success in the interaction. After some rounds of negotiation, a shared set of form-meaning mappings emerges without the need for central control. Similar experiments using neural networks were reported by Batali (1998).

The notion of a **language game** was first introduced by Steels (1996c, also see Steels, 2001a, for an introduction). Language games are routinised local interactions or scripts. An example of a language game in natural languages is a speaker who asks *Can you pass me the salt, please?*. The language game is a success if the hearer passes the salt or even if she responds that she refuses to do so (but at least she understood the message). The game is a failure if the hearer passes the pepper instead of the salt or if she just shrugs her shoulders. The speaker can then point to the salt, which gives the hearer some additional clues as to what she meant with *salt*. An iteration of language games is called a **dialogue**, but this more complex interaction pattern has not been studied anymore since Steels (1996c).

In these first exeriments, the meaning space of the agents was predefined. However, since no concepts or categories are assumed to be innate in this line of work, several experiments have been conducted investigating how agents can create their own concepts and meanings. The first breakthrough was reported in Steels (1996b), in which agents created perceptual categories through **discrimination games**. In a discrimination game, an agent tries to discriminate a certain object from the other objects in the context by creating or using a set of one or more distinctive features for that object. In the next breakthrough experiment, discrimination games and language games were coupled to each other so that agents did not only self-organise a lexicon, but also used the lexicon for sharing concepts or perceptual distinctions (Steels, 1997a,c, 1998a, 1999a,b). In a next step, it was shown how these ideas can be grounded in actual robotic agents (Steels, 2001b, 2002a; Steels & Vogt, 1997; Vogt, 2000). The structural coupling of concepts and lexicons has also been successfully applied to the domains of colour (Belpaeme, 2002; Belpaeme & Bleys, 2005; Steels & Belpaeme, 2005) and space (Loetzsch *et al.*, 2008; Steels & Loetzsch, 2008).

The research on lexicon formation steadily grew and touched upon topics such as how lexicons can continue to change and evolve because of language contact and population dynamics (Steels, 1997b; Steels & McIntyre, 1999) and stochasticity in cultural transmission (Kaplan *et al.*, 1998; Steels & Kaplan, 1998a,b). The experimenters also developed the notion of a **semiotic landscape**, a powerful framework to study the **semiotic dynamics** involved in the language games (Steels & Kaplan, 1999a,b). The research culminated in the Talking Heads experiment which involved thousands of agents travelling over the internet in order to play language games with each other (Steels, 1999c, 2000c, 2001a; Steels & Kaplan, 1999a,b, 2002; Steels *et al.*, 2002).

The research on lexicon formation is still being pursued today. For example, the Naming Game, which first appeared in Steels (1997e), has recently been implemented in humanoid robots that autonomously have to recognise objects as individuals and then agree on names for them. The Naming Game has also been picked up by scientists from statistical physics and complex systems who search for scaling laws and who investigate the long-term behaviour of the system using mathematical models (Baronchelli *et al.*, 2006; De Vylder, 2007). Other recent work using computational modeling investigates how word meanings can be more flexible and how the formation of lexicons can scale up to much larger worlds (Wellens *et al.*, 2008).

3.3.2 Towards grammar

Even though the first decade has been largely spent on investigating the properties of emergent adaptive lexicons, the formation of grammar has always been on the research agenda with first attempts as early as 1997 (Steels, 1997d). The research strategy involves moving all the insights gained from the experiments on lexical languages to the domain of grammar and identify which additional mechanisms and ideas are needed for the formation of languages featuring grammatical properties (Steels, 2005a).

The first steps towards grammar involved a pregrammatical stage of **multiple word utterances**, which was first investigated by Steels (1996a). In this experiment, multiple word utterances emerge naturally as the set of distinctive features for talking about objects expands and the agents have to adapt to cope with it. Van Looveren (2005) then showed how a multiword naming game can yield a more efficient communication system because a smaller lexicon could be used for naming objects. However, none of these experiments involved any grammar.

At the end of the nineties, a significant breakthrough was achieved which resulted in the general roadmap for investigating the emergence and formation of grammar that is still being followed today (Steels, 1999c, p. 44–47; Steels, 2000a). Whereas lexical languages are perfectly suited for language games involving only one object, grammar becomes useful when agents have the possibility of communicating about multiple objects because the search space becomes exponentially larger. Grammar thus emerges not in order to reduce inventory size or to become more learnable (as is proposed by genetic and Iterated Learning Models, see section 3.1), but rather to **reduce the complexity of semantic parsing** for the hearer (this idea has been formalised and operationalised by Steels, 2005b, also see section 3.1.3 and 5.2). Luc Steels then worked on systems for studying the formation of **compositional meanings** (see section 3.3.3) and for the formation of grammar to take care of these compositional meanings.

Van Looveren (2005) applied these ideas to his experiments on multiple word games and lifted the agents' assumption that multiple words always refer to the same object. For example, the utterance *yellow ball* might refer to one object (a yellow ball) or at least two objects (some yellow thing and a ball). When faced with this referential uncertainty, the agents can exploit a simple syntax for indicating to the hearer which words refer to the same object. referential uncertainty has also been investigated from the viewpoint of pattern formation (as another pregrammatical stage, Steels *et al.*, 2007) and as a trigger for introducing additional syntax (Steels & Wellens, 2006).

Another key issue in the formation of grammar is the question of how agents can autonomously detect when additional constraints or grammar might become useful in order to improve their communicative success. The answer is ‘re-entrance’ (Steels, 2003b), a strategy which was already present in the experiments on the formation of lexicons but which had to be developed further to fit into a grammatical framework. Re-entrance can be seen as some kind of self-monitoring in which the speaker first simulates the effect of her utterance on the hearer by taking herself as a model. If she detects problems such as ambiguity or an explosion of the search space during parsing, she will adapt her linguistic behaviour by adding more constraints or choosing a different verbalisation. Similarly, the hearer can perform re-entrance for learning novelties in the speaker’s utterance (for more details, see Chapter 6). A similar mechanism called ‘the obverter strategy’ can be found in other negotiation-based models as well (Smith, 2003a). Another way to increase the autonomy of the agents is to offer them strategies for self-assessing what kind of communicative goals they can attain given their present linguistic experience (called the ‘autotelic principle’, Steels, 2004a,b; Steels & Wellens, 2007).

Most of the above ideas were put to practice in 2001 in the first ‘case experiment’, which I will describe in Chapter 6 and of which some results were published in Steels (2004c) and to a lesser extent in Steels (2003b, 2007). Luc Steels implemented a unification-based grammar formalism to support the experiment, which ultimately led to the first design of Fluid Construction Grammar (FCG, also see Chapter 5). FCG is under constant development to meet the demands and requirements of new experiments such as the following:

- A bidirectional and uniform way of language processing (Steels & De Beule, 2006). This feature is needed for allowing agents to act both as a speaker and as a hearer; and for allowing the agents to perform re-entrance;
- A way to deal with compositional semantics and to link meanings to each other (Steels *et al.*, 2005);
- A way to support hierarchical structure building operations both in production and parsing (De Beule & Steels, 2005).

The name ‘Fluid Construction Grammar’ comes from the fact that FCG has many features in common with (vanilla) construction grammar (Croft, 2005) and that it aims at investigating the ‘fluidity’ of language formation (i.e. various degrees of entrenchment of linguistic units). A software implementation of FCG, incorporated in a more general cognitive architecture called ‘Babel2’, has been made freely available at *www.emergent-languages.org*.

Fluid Construction Grammar has already been used for investigating the formation of compositionality (De Beule & Bergen, 2006), recursion and hierarchy (Bleys, 2008; De Beule, 2007, 2008), structures for expressing second-order meanings (Steels & Bleys, 2005) and semantic roles (Steels, 2004c; van Trijp, 2008b). In part II of this thesis; I will report experiments on the formation of a case grammar in FCG.

3.3.3 Other research avenues

The above account of the history of the research on language as a complex adaptive system did not refer to the experiments on the emergence of vowel systems and phonology (de Boer, 2000; Oudeyer, 2005). I also left out the work performed on event recognition, but I will come back to this in Chapter 6. Another area that I left largely uncovered was the research into conceptualisation and the formation of complex meanings.

Together with the key insights on the triggers and functions of grammar at the end of the ninetees, Steels developed a constraint-based system for studying the emergence of compositional meaning (Steels, 2000a) which uses constraint propagation both for conceptualisation and interpretation. In the system, a set of constraints can be composed into some kind of semantic program which the speaker wants the hearer to perform. For example, if the speaker wants to draw the hearer's attention to a particular ball in the context, she has to conceptualise a network of meanings or constraints that will help the hearer to correctly identify this ball. For example, the utterance *the red ball* indicates to the hearer that she has to (a) filter the objects in the context and retain those that match with the prototype [BALL], (b) filter this set of balls and retain the red ones, and (c) pick out the one remaining object (its uniqueness was indicated by the determiner *the* in combination with the singular form *ball*). The system allows for the composition of second-order semantics (e.g. *the bigger ball*) and context-sensitive meanings (e.g. a 'small' elephant is still bigger than a 'big' mouse).

Even though at that time there was already an operational system, the research on compositional meanings was put on a hold to first develop a grammar formalism that could support it. With the recent advances in Fluid Construction Grammar, the research got picked up again and first experiments have already been reported that couple meanings produced by the system to FCG and vice versa (Bleys, 2008; Steels & Bleys, 2005). The system has also been completely re-implemented and improved, and is now ready for use (see Van den Broeck, 2007, 2008, for an introduction).

3.4 Summary: modeling the formation of grammar

In this chapter I gave a quick overview of the main approaches to modeling the formation of language and indicated how they relate to the linguistic theories discussed in Chapter 2. I showed that each approach assumes a selectionist system to explain variation and language change, but that they have different opinions of what the replicating units are in language. Genetic models assume natural selection at the level of the genes, Iterated Learning Models assume cultural selection at the transmission of a language from one generation to the next, and problem-solving models assume cultural selection in each communicative interaction. Whereas genetic and Iterated Learning Models propose that agents use their languages (mostly) unchanged throughout the rest of their lives, the problem-solving model claims that agents continuously shape and reshape their knowledge about their language. Another point of difference is what forces are hypothesised to trigger the emergence of grammar. Genetic models focus on inventory optimisation, Iterated Learning Models focus on transmission bottlenecks and the Poverty of the Stimulus, and problem-solving models focus on how grammar can be exploited for reducing the complexity of semantic parsing in order to optimise communicative success.

I illustrated the three above approaches through the debate on holistic versus compositional languages and argued that the problem-solving approach obtained significantly better results. Since I also follow this approach in this thesis, I gave a quick overview of what assessment criteria are important for a sound application of the problem-solving model. I indicated why taking the design stance towards the formation of language is useful, but also pointed to the limitations of the methodology. An interdisciplinary dialogue is therefore of utmost importance for the relevance of the model. The work in this thesis in the first place aims at a dialogue with linguistic theory.

Finally, I gave a short overview of the body of research that has been performed in the problem-solving approach. I pointed to the relevant background literature and key ideas that have emerged in more than a decade of research. The work in this thesis is part of moving the research into the domain of grammar and is heavily indebted to the prior work of the past thirteen years.

Chapter 4

A grammar for marking event structure

The more than 6.000 languages in the world have each in their unique way developed a grammar which allows for an astonishing range of expressiveness. We can just as easily engage in hair-splitting philosophical debates as we can start gossiping about as soon as somebody leaves the room. We can talk about the first landing on the moon even though it happened decades ago, we can entertain our young ones with stories even though they're not always real, and we can hurt and offend others even though we did not touch them. Luckily, we can also whisper soothing words into people's ears in order to comfort them: there really seem to be no limits to what we can express through language.

At the heart of the grammar that allows us to communicate all these complex meanings lies the way that relations between words are organised. If you want to hear about juicy facts such as who kissed whom, grammar can help you finding out using various strategies such as word order, case marking and verb agreement. How such a grammar for marking argument structure (in the form of case markers) can be formed is the main topic of this thesis.

In this Chapter, I will therefore apply the assumptions of Chapter 2 to the formation and evolution of case marking systems. First of all, it is important to have an appreciation of the complexity of the problem that 'argument realisation' poses us. I will deal with this in section 4.1 by reviewing three approaches to argument realisation. This discussion will form the basis for my own proposal for argument structure in Fluid Construction Grammar, which I will present in Chapter 5. In section 4.2 I will illustrate the life cycle of a case marker supported with evidence from natural languages. This whole chapter aims at (a) defining a guidance and reference point for understanding the experiments in Part II, and (b) making the exact scope and more specific theoretical foundations of the experiments more clear.

4.1 The problem of argument realisation

Most linguists agree that there is a strong connection between the semantic representation of a verb and its morphosyntactic behaviour. Unfortunately, the exact nature of this connection – which is better known as the problem of 'argument realisation' – is still a largely unresolved issue in linguistics. This section offers a brief survey of the most influential theoretical accounts of argument realisation and points to their shortcomings. A more complete overview and thorough discussion can be found in Levin & Rappaport Hovav (2005).

4.1.1 The semantic role list

The notion of 'semantic roles' (also called case-, teta-, argument- or thematic roles) has a longstanding history in linguistics which dates back from Panini's grammar of Sanskrit, but only resurfaced in modern linguistics under the influence of Gruber (1976, originally written in 1965) and especially Fillmore (1968). As explained in section 2.4.1, most syntactic theories of that time were trying to come up with a set of syntactic rules which could generate all possible (and well-formed) sentences of a language. As examples 2.4 and 2.5 showed, however, semantic constraints soon became necessary to further limit the 'generative power' of these syntactic rules. The examples are reprinted here:

(4.1) I broke his leg. / *I broke him on the leg.

(4.2) I hit his leg. / I hit him on the leg.
(Fillmore, 1970, p. 126, quoted from Levin & Rappaport Hovav, 2005, p. 2)

Fillmore (1968) proposed a small list of semantic roles ('case roles' in his terminology) such as 'agent', 'goal', 'source', 'instrument', etc. A subset of these semantic roles is then specified in the lexical entry of a verb and makes up the verb's 'case-frame' (also called 'predicate-frame' or 'theta-grid'). For example, a verb like *break* can be listed as a two-place predicate assigning the roles of 'agent' and 'patient' to its arguments.

Similar approaches have been widely adopted in various linguistic theories, both generative and functional. Croft (1991, p. 156) writes that these varieties all have the following three assumptions in common:

> i. Thematic (case) roles are defined as semantic primitives (i.e., semantically unanalyzable).
> ii. Thematic roles are defined independent of the semantics of the verb, which is also left unanalyzed (primitive).
> iii. There are only a small finite number of thematic roles.

Croft argues that these three assumptions lead to a reductionist approach in which the number of semantic roles is minimised as much as possible in order to limit the number of grammatical rules that operate on them. This reductionist approach however suffers from a number of critical fallacies when trying to account for empirical data.

Problems with grain-size. One problem has to do with finding the right 'grain-size' in the definition of semantic roles (Levin & Rappaport Hovav, 2005, p. 39). Every argument structure generalisation that can be made at first sight proves to be inadequate once a more detailed analysis is done. Levin & Rappaport Hovav illustrate this through an argument structure rule suggested in their earlier work (Levin & Rappaport Hovav, 1995, p. 39) in which the 'immediate cause' of an event is realised as the subject of an English sentence. 'Immediate cause' is a broader category than 'agent' because also non-agentive entities can cause an event. However, as the following examples show, English seems to make a distinction between different subclasses of 'immediate cause' in certain constructions:

(4.3) a. His uncle died from / of pneumonia.
b. Pneumonia killed his uncle.

(4.4) a. Brutus killed Caesar.
b. *Caesar died from / of Brutus.
(Levin & Rappaport Hovav, 2005, p. 40, examples 12-13)

Faced with the empirical data, many linguists have therefore abandoned the idea of a fixed number of semantic roles. For example, Palmer (1994) distinguishes between a generalised role 'Agent' (with a capital A) and the notional role 'agent' (the performer of an action). Similarly, Givón (2001) states that semantic roles such as 'agent' are prototypes which have various subclasses. It is also telling that so far, no consensus has been reached over determining the exact number and nature of the semantic roles. In fact, as Croft (1991) argues: *"taken to its logical conclusion, a fine-grained analysis of thematic roles will result in a unique case-frame for almost every verb in a natural language"* (p. 158). The data thus suggest that the third assumption stated above is false: there is no reason to believe that the number of semantic roles is very small.

The influence of verb meaning. Many of the problems of the definition of semantic roles can be solved if the second assumption is given up: semantic roles cannot be investigated as abstract entities apart from the lexical meanings of verbs. Examples 4.1 and 4.2 are good illustrations of how the lexical entailment of a verb influences its distribution patterns: *break* entails that the patient undergoes a change of state, whereas *hit* does not share this entailment. Another well-known example is the difference between the verbs *murder* and *kill*: *murder* expects a volitional agent whereas *kill* can involve virtually any cause of death. These observations have led many researchers to investigate event structure, a solution which will be discussed in section 4.1.2.

If the definition of semantic roles is closely related to the semantics of verbs, then the first assumption may be false as well. Semantic roles should not be treated as primitive or atomic categories, but as categories that either have to be derived from more basic primitives, or that have to be constructed and learned. As non-primitive (and arguably non-universal) categories, we might expect a lot of cross-linguistic variation and this is indeed suggested by many typological studies (e.g. Croft, 1991, 2001; Dryer, 1997). Haspelmath (2007) writes that (for grammar in general) it seems impossible to find a set of innate categories, *"[o]n the contrary, almost every newly described language presents us with some "crazy" new category that hardly fits existing taxonomies"* (p. 119).

Relations between semantic roles. The approach of a fixed list of semantic roles also has the problem that no relations are defined between the roles. This makes it impossible to explain why certain combinations of semantic roles could never occur, or it may lead to errors in descriptions. For example, Davidse (1996) shows that the various functions of the English dative in the alternation between the ditransitive construction (such as *I gave him a book*) and its prepositional counterpart (such as *I gave the book to him*) can only be understood if the dative is investigated as a part of these constructions rather than as *"an isolated element of structure"* (p. 291). Instead of the traditional approach in which the dative is described as some kind of 'possessor' or 'recipient', the whole ditransitive pattern reveals various relations concerning *caused* possession, involving a causer, some object (potentially) being transferred, and a (potential) receiver. This analysis reveals a multifunctional grammatical paradigm containing at least eight different senses for the English ditransitive, which can be seen as a *"grammatically-construed [...] form of polysemy"* (p. 333). Similarly, Goldberg (1995, p. 31–38) writes that constructions are polysemous and she lists at least six different senses for the English ditransitive construction (see examples 4.25 – 4.30).

Multiple argument realisation. Finally, the lexicalist approach fails to account for 'multiple argument realisation', which means that verbs can occur in several argument structures. This phenomenon has for a long time been regarded as a rather exceptional fact due to polysemy of a verb: for each different argument structure, a different lexical entry or a different case-frame was posited in the lexicon. However, multiple argument realisation seems to be pervasive in language as can be seen in the following examples:

(4.5) a. Terry swept.
b. Terry swept the floor.
c. Terry swept the leaves into the corner.
d. Terry swept the floor clean.
(Levin & Rappaport Hovav, 2005, p. 188, example 9)

The problem of multiple argument realisation has especially been used by construction grammarians to criticise the lexical approach. I will go deeper into the constructivist alternative in section 4.1.3.

4.1.2 Lexical decomposition and event structure

The observation that the definition of semantic roles *"must mesh in some natural way with the lexical semantics of the verbs that govern them"* (Croft, 1991, p. 158) demands for a richer representation of these semantics. Many contemporary theories

Verb Class	Logical Structure [LS]
STATE	**predicate´ (x)** or **(x,** y)
ACTIVITY	**do´ (x, [predicate´ (x)** or **(x,** y)])
ACHIEVEMENT	INGR **predicate´ (x)** or **(x,** y), or
	INGR **do´ (x, [predicate´ (x)** or **(x,** y)])
ACCOMPLISHMENT	BECOME **predicate´ (x)** or **(x,** y), or
	BECOME **do´ (x, [predicate´ (x)** or **(x,** y)])
ACTIVE ACCOMPLISHMENT	**do´ (x, [predicate$_1$´ (x)** or **(x,** y)])
	BECOME **predicate$_1$´** (z, **(x,**) or (y)
CAUSATIVE	α CAUSE β, where α, β are LSs of any type

Table 4.1: The lexical representation for classes of verbs in Role & Reference Grammar (Van Valin, 2004, table 1).

have therefore adopted an ‘event structure’ approach or include ‘lexical decomposition’ representations. Both labels are often used as synonyms for each other in the literature, but outspoken ‘lexical decomposition’ theories (most of them influenced by Dowty, 1979) focus more on the possible entailments of a predicate or event, whereas ‘event structure’ representations focus on the decomposition of the event itself. I will briefly discuss one example of both in this section.

Lexical decomposition. One example of lexical decomposition is found in Role & Reference Grammar (RRG, Van Valin, 1999, 2004). RRG hypothesises that verbs can be categorised according to their aspect into the classes of state, activity, achievement, accomplishment, active accomplishment and causative (see Table 4.1). Examples are (Van Valin, 2004, taken from example 3):

(4.6) STATES:

Leon is a fool.	**be´** (Leon, [**fool’**])
The window is shattered.	**shattered´** (window)

(4.7) ACTIVITIES:

The children cried.	**do´** (children, [**cry´** (children)])
Carl ate snails.	**do´** (Carl, [**eat´** (Carl, snails)])

(4.8) ACHIEVEMENTS:

The window shattered.	INGR **shattered´** (window)
John glimpsed the picture.	INGR **see´** (John, picture)

(4.9) ACCOMPLISHMENTS:

The snow melted.	BECOME **melted´** (snow)
Mary learned French.	BECOME **know´** (Mary, French)

Arg of DO	1st arg of **do´** (x, ...	1st arg of **pred´** (x,y)	2nd arg of **pred´** (x,y)	Arg of state **pred´** (x)
AGENT	EFFECTOR	LOCATION	THEME	PATIENT
	MOVER	PERCEIVER	STIMULUS	ENTITY
	PERFORMER	COGNIZER	CONTENT	
	CONSUMER	WANTER	DESIRE	
	CREATOR	JUDGER	JUDGMENT	
	SPEAKER	POSSESSOR	POSSESSED	
	OBSERVER	EXPERIENCER	TARGET	
	USER	EMOTER	ATTRIBUTE	
	...	ATTRIBUANT	...	

Figure 4.1: All the semantic roles are placed in a thematic hierarchy in Role & Reference Grammar going from ‘prototypical agent’ to ‘prototypical patient’. The generalised semantic macro-roles of RRG (Actor and Undergoer) select their arguments based on this scale (Van Valin, 2004, adapted from figure 2).

(4.10) ACTIVE ACCOMPLISHMENTS:

Carl ate the snail.	**do´** (Carl, [**eat´** (Carl, snail)]) & BECOME **eaten´** (snail)
Paul ran to the store.	**do´** (Paul, [**run´** (Paul)]) & BECOME **be-at´** (store, Paul)

(4.11) CAUSATIVES:

The dog scared the boy.	[**do´** (dog, ∅)] CAUSE [**feel´** (boy, [**afraid´**])]
Max broke the window.	[**do´** (Max, ∅)] CAUSE [BECOME **broken´** (window)]

Semantic roles can now be defined in terms of the argument positions in the decomposed logical structure. For example, in a state with a single argument such as *the window is broken*, decomposed as **broken´ (window)**, ‘window’ is the ‘patient’. Even though there is a large number of semantic relations in RRG, they all fall into the five aforementioned classes and form a continuum from ‘prototypical agent’ to ‘prototypical patient’ (see Figure 4.1).

This continuum, also known in the literature as a ‘thematic hierarchy’, provides a ranking of semantic roles, which is then used by linking rules in order to decide which role becomes subject and which role becomes object. The first proposal of such a thematic hierarchy can be found in Fillmore (1968, p. 33) to account for examples such as the following:

(4.12) a. I opened the door with the key.

b. The key opened the door.

c. The door opened.
(Croft, 1998, p. 27, example 5; examples adapted from Fillmore, 1968)

The basic idea is that the semantic role which is placed highest in the thematic hierarchy becomes subject. In example 4.12.a. the ‘agent’ becomes subject because it is ranked higher than the ‘patient’ or the ‘instrument’. The other two sentences do not have an agent so an other semantic role fills the subject position (also see Levin & Rappaport Hovav, 2005, chapter 6, for a critical review of thematic hierarchies).

On top of the thematic hierarchy, RRG proposes that languages feature two generalised semantic ‘macro-roles’ called ‘Actor’ and ‘Undergoer’. These macro-roles are mainly motivated by the (grammatical) importance of the agent-patient distinction in so many languages (however, see section 4.2.2 and Palmer, 1994, for arguments against a universal distinction of agent versus patient). The Actor role is placed on the left of the thematic hierarchy (see Figure 4.1) and is opposed to the Undergoer role which is placed on the right.

The argument linking system in RRG thus starts with the event-specific participant roles (e.g. a runner, a giver, etc.) which can be mapped onto semantic roles depending on the logical structure of the verb’s semantics (more recent versions of RRG, however, deny theoretical status to semantic roles; see Van Valin, 1999). These semantic roles, which are argument positions in the decomposed logical structure, are ranked according to a thematic hierarchy going from Actor to Undergoer, which are both semantic macro-roles. The macro-roles themselves are then mapped onto syntactic structure. RRG hypothesises that all levels up to the macro-roles are universal, whereas the mapping from macro-roles to syntactic functions is language-specific. Accusative languages like English take by default the Actor as the subject, whereas ergative languages like Dyirbal choose the Undergoer as the default subject.

The Role & Reference Grammar approach of lexical decomposition shares some crucial theoretical constructs with other proposals such as Dowty (1991); Jackendoff (1990); and Pinker (1989). According to Croft (1998, p. 21–22), these are:

- Event structure: all of these theories assume that argument linking should be based on a decompositional representation of the causal/aspectual structure of events;
- Thematic hierarchy: semantic roles are not treated as semantic primitives anymore, but are derived from event structure. With the exclusion of Dowty (1991), semantic roles are ranked in a thematic hierarchy;
- Macro-roles: thematic hierarchies may be appealing to many linguists, but they are highly problematic and insufficient in determining argument linking (see

below). Most theories therefore assume some additional layer of macro-roles. These are either semantically defined, as in RRG and the protoroles in Dowty (1991), but other theories place them on the syntactic structure such as A-marking in Lexical Conceptual Structure (Jackendoff, 1990, p. 254).

- Linking rules: all theories posit rules that specify how the participants of an event should be linked to syntactic structure.

Lexical decomposition (in its narrow sense as I defined it in the beginning of this section) suffers some of the same problems as the semantic role lists because their architecture remains more or less the same: semantic roles are derived from decomposed lexical semantics, but once this is achieved, the rest of the event structure information is discarded. This has the following consequences (see Croft, 1998, p. 27–41 for a more in-depth review):

- Most approaches take a reductionist approach to lexical decomposition and look for 'universal' decompositions (e.g. the five classes in RRG). However, no consensus has been reached about the number and nature of these semantic primitives (Nuyts, 1992) so the aforementioned problem of grain-size in the definition of semantic roles has merely shifted to a different level of description;

- Semantic roles are defined as slots in the event structure or lexical decomposition. However, as argued before, a semantic role only makes sense within a certain pattern and not as an isolated category. A 'goal' can only be a goal if some entity is moving towards it. So (most) theories of lexical decomposition fail to explain the relations between semantic roles;

- Thematic hierarchies are needed to explain for example why the semantic 'agent' typically maps onto the syntactic 'subject' (at least in accusative languages), but again the same problems of semantic role lists reoccur. First of all, a ranking of semantic roles does not explain the relations between roles (which is necessary for the reasons argued before). Second, due to discussions about the number of semantic roles and cross-linguistic differences, there is no consensus whatsoever about what the 'universal' thematic hierarchy should be (compare for example the hierarchies in Dik, 1997; Fillmore, 1968; Givón, 2001; Jackendoff, 1990; Van Valin, 2004);

- The additional layer of macro-roles (whether syntactic or semantic) has to be posited in the lexical entry of the verb because there may be exceptions or marked links which overrule the default assignment according to the hierarchy. As such the macro-roles replace the semantic role list without solving its problems.

Event structure. Croft (1998) argues that a richer representation of event structure can eliminate the necessity of thematic hierarchies and macro-roles. Next to the causal-aspectual decomposition of a verb, force-dynamics and the verbal profile are crucial in argument linking. Croft proposes that force-dynamic relations among participants (i.e. how participants of an event relate to each other in a causal chain) are more critical to argument linking than semantic roles so there is no need for ranking semantic roles in a thematic hierarchy. With 'verbal profile', Croft means that the same force-dynamic conceptual semantic structure may lead to different verbs which each focus on various segments of the causal chain. Argument linking then occurs according to four universal (but not global; see below) linking rules ('>' stands for 'precedes in the force-dynamic chain'):

1. The verbal profile is delimited by Subject and Object (if any [...])
2. Subject > Object
3. Antecedent Oblique > Object > Subsequent Oblique
4. Subject > Incorporated Noun > Object (if any)

(Croft, 1998, p. 24)

Croft's use of the word 'universal' should not be confused with how it is used in generative grammar. Rather than interpreting 'subject' and 'object' as innate linguistic categories (in fact, Croft radically rejects the existence of such categories, see Croft, 2001), he takes a typological-functionalist approach in which *"universal grammar consists of a variety of functions that language is intended to serve"* (Croft, 1991, p. 1). This approach allows him to take a 'polysemy' or 'usage type' approach in defining grammatical categories rather than the so-called 'Gesamtbedeutung' (i.e. one single abstract meaning or definition) for categories (ibid., at p. 183; also see Blake, 1994, p. 36–47 for more on this). In other words, linking rules 2–4 should be seen as implicational universals applied to case marking and to a lesser extent to other coding constructions such as word order or agreement (Croft, 1991, chapter 5).

There are still other tricky issues in interpreting the 'universal linking rules'. A first complexity lies in the distinction that Croft makes between conceptual representation and construal operations on the one hand (which are hypothesised to be universal), and semantic representation on the other (which is a conventional and language-specific construal of conceptual representation). Conceptualisation of events is thus language-specific and may even alter conceptual representation. The second complexity of the linking rules is that they are not global: they do not apply to all event types. They only apply to those events which have clear force-dynamics. Given this hypothesis argument linking should be consistent across languages for events with clear force-dynamics, but vary when this is not the case.

To illustrate Croft's approach, we can take a look again at the sentences in example 4.12, which have 'clear' force-dynamics underlying them. The agent in 4.12.a. acts on a patient (the door) and precedes it in the causal chain. So the agent becomes the subject and the patient becomes the object. When there is no agent but an instrument, such as in 4.12.b., the instrument becomes subject because it also acts upon the patient and hence precedes it in the causal chain. Finally, if there is no instrument, the patient becomes the subject. Including the force-dynamics of events thus makes a thematic hierarchy unnecessary. For events that do not have clear force-dynamics (e.g. mental state verbs), linking depends on the construal of the mental event.

Argument linking is also dependent on the verbal profile, which is decided by an interplay between the verb and the constructions in which it occurs (see the last paragraphs of section 4.1.3 for more details on the meaning of 'verbal profile'). Croft gives the example of *break* versus *broken*. *Break* denotes a process whereas *broken* denotes a resulting state (after a break-event). Both words thus specify the same event type, but have a different profile. The verb profile is independent of force-dynamics so alternative argument linkings are possible (as in example 4.12).

Croft uses a geometric representation for the semantic representation of event structure, which is illustrated for a locative alternation in examples 4.13 and 4.14 (Croft, 1998, p. 39, examples 16–17). The first row in this representation indicates the force-dynamic relations in a causal chain between participants. '···' stands for a force-dynamic relationship (cause), '===' for a process, and ' —' for a state. In both examples, the same causal chain or event structure is underlying the analysis. The difference then lies in the verbal profile: in example 4.13, the verb only profiles x causing y to undergo a process leading to a resulting state (y) (indicated by '***'). '●●●' stands for an adposition, which profiles an additional part of the event. In example 4.14, the verb profiles the whole event structure. The resulting state of the wall (z) is linked to object position, and the paint is marked as an antecedent oblique (linking rule 3).

(4.13) Jane sprayed paint on the wall.

x ······ y ====== (y) ——— z
****** *spray* *************●●●*on* ●●●●●●●
Sbj Obj S.Obl
Jane *paint* *wall*

(4.14) Jane sprayed the wall with paint.

x ······ y ====== (y) ——— z
*** *spray* *****************************
Sbj A. Obl Obj
Jane *paint* *wall*

Remaining challenges. Croft's causal/aspectual approach to event structure makes a strong attempt to 'go back to basics' because it eliminates the need for theoretical constructs such as thematic hierarchies and macro-roles. This can be achieved by allowing a richer representation of event structure and a focus on the relations between participants in a causal chain. Croft does not claim that there is a universal set of semantic primitives that make up event structure, but that argument linking constructions only need to make reference to the (construal of) force-dynamics of events. Even though Croft (1991, 1998, in prep.) applies to a broad range of empirical data, some serious issues remain.

The main problem is how Croft separates the problem of argument linking from how arguments are realised morpho-syntactically. In his approach, semantic roles do not matter for argument linking:

> A force-dynamic analysis specifies the transmission of force relationships between participants in events. No analysis into thematic [i.e. semantic] roles, either a small broadly defined set or a large fine-grained set, is necessary: all that one must know is the force-dynamic relationships among participants. In other words, thematic role types is the wrong generalization across participant roles in different event types; force-dynamic relations is the right generalization. (Croft, 1998, p. 31)

Let us examine this hypothesis more closely by looking at examples 4.13 and 4.14. Both examples feature the same semantic representation for event structure, but this doesn't fit the right interpretation of both sentences: in *Jane sprayed paint on the wall*, there is no implication that the wall is fully covered with paint. In *Jane sprayed the wall with paint*, however, the speaker *does* indicate that the wall is completely covered. A correct semantic representation would thus involve a resulting state for 'z' (the wall) and not for y:

(4.15) Jane sprayed the wall with paint.

x ······ y ······ z ====== (z)
*** *spray* *****************************

Sbj	A. Obl	Obj
Jane	*paint*	*wall*

However, this analysis points to another problem: the oblique in example 4.15 (*with paint*) is analysed as profiled by the verb, whereas the oblique in 4.13 is profiled by an adposition. Is there evidence that 'antecedent obliques' are always profiled by the verb and 'subsequent obliques' always by adpositions? With linking rule 3, Croft (1991) points to a strong pattern in language: if a language has different oblique case markers, they will (usually) only mark either antecedent (e.g. *by, from*) or subsequent obliques (e.g. *to, on*) of a causal chain (Croft, 1991, chapter 5). However, as Croft acknowledges himself, the distinction breaks down if there are no different markers and not all cases are as clear-cut. An alternative explanation for the distinction is that oblique case markers group together semantically related roles (see section 4.2.4): *to*, for example, groups together the 'goal' of a motion but also the recipient in a ditransitive (which can also be seen as the metaphorical goal of a transfer of possession).

A third issue with the force-dynamics-only analysis is that there seem to be hardly any events which have or which *only* involve sub-events with clear force-dynamics. Even a light verb such as *give* contains at least the non-force-dynamic notion of 'possession' so it has to be construed. Since the construal of events is highly dependent on semantic meaning, linguistic perspective or information structure, linking patterns are equally dependent. This can be illustrated through the following examples:

(4.16) The stone broke the window.

(4.17) He received a book (from her / *by her).

(4.18) He was given a book (*from her / by her).

In Croft's approach, the verb *broke* in example 4.16 only profiles the instrument and the patient of a break-event. Since the instrument antecedes the patient in the causal chain (and the agent is not profiled), it is linked to subject position. However, the sentence *the stone broke the window* involves more than merely stating that the stone antecedes the breaking: the speaker here presents the event as if the stone was the cause (and not the instrument) of the break-event. The construal of (the causal chain of) a break-event is thus dependent on semantics. That a break-event does not always have clear force-dynamics seems to be confirmed by the fact that instruments cannot take subject position in every language (Croft, 1991, p. 152). Moreover, the unacceptability of utterances such as **he broke* (meaning that he broke something) can only be explained in terms of semantics, because there seems to be no reason why a verb cannot uniquely profile the cause/agent of an event (e.g. *he swept*). The utterance *he broke* can only be understood as 'he psychologically broke (under the pressure)' in which the subject is the undergoer of the break-event.

Also consider example 4.17, which seemingly violates the linking rules of a causal chain. Croft (in prep., chapter 5) analyses a receive-event as a causal chain of a non-profiled donor which acts on a recipient (that gains possession of something) and which results in a gift (which now has shifted in ownership). This construal is dependent on linguistic perspective: *receive* profiles the scene from the viewpoint of the receiver rather than the giver. However, this profile is exactly the same as how Croft (1998, p. 55–56) deals with passives: in 4.18, the donor is not profiled and becomes an antecedent oblique marked with *by*. The donor in a receive-event, however, cannot be marked with *by* but as a source (*from*). The receiver is in fact conceptualised as an agent-like participant who is 'capable' of receiving, as is clearly the case in *the book was received well by the critics*.

The evidence suggests that if argument linking were completely separated from grammatical distinctions (and would only depend on the causal/aspectual chain), argument realisation would involve the following three steps: (1) speakers would create construals for specific events dependent on the semantic conventions in their speech community, (2) linking is then done based on these construals, and (3) semantic roles would then be derived in order to determine the final morpho-syntactic realisation of arguments. As I will show in the next chapter, we can collapse these steps into one if we involve semantic roles again as a crucial part of argument linking.

In this section I did not go into much details regarding Croft's definition of a 'verbal profile' or his position within linguistic theory. This will become much clearer in the following section in which I explain the constructivist approach to argument realisation. Croft's proposal can be seen as constructivist (also see Croft, 2001), but more nuanced than e.g. Goldberg (1995), which perhaps still offers the most explicit construction grammar approach to argument structure to date.

4.1.3 A constructivist approach

A current topic in the question of argument realisation is the debate between traditional 'lexicalist' and the more recent 'constructivist' approaches. The lexicalist account, among which the semantic role list theories and most of the lexical decomposition proposals, specify a verb's argument structures in its lexical entry and make use of 'lexical rules' to project this argument structure on the syntactic structure of the sentence (for an overview of the various existing mapping rules, see Levin & Rappaport Hovav, 2005, chapter 5). The constructivist account argues that there is no sharp distinction between the lexicon and syntax and that even argument structure constructions are symbolic units (i.e. conventionalised form-meaning mappings), so some parts of the meaning should be attributed to the argument structure constructions instead of to the verb. The contrast can be illustrated through the following examples:

(4.19) She baked a cake.

(4.20) She baked him a cake.

Both sentences feature the verb *bake* but are slightly different in meaning: example 4.19 only means that someone baked a cake, whereas example 4.20 means that the cake was created with the intention of giving it to someone. In the lexicalist approach, this would mean that there are either (a) two different lexical entries for the verb *bake* with their own case-frame, (b) that the lexical entry of *bake* has two distinct case-frames, or (c) that there is a lexical rule which derives the meaning of *bake* in example 4.20 from the meaning of *bake* in example 4.19. Examples of lexicalist approaches are Bresnan (1982); Dik (1997); Fillmore (1968); and Pinker (1989). Constructivists, on the other hand, argue that in example 4.20, the meaning of 'beneficiary' is added by the construction itself and not by the verb. The so-called ditransitive construction is argued to map various meanings of caused possession to the ditransitive pattern (Goldberg, 1995, chapter 6).

Arguments against the lexicalist approach. The lexicalist view has been dominant in linguistics for decades, but has recently received some serious criticism from cognitive linguistics. I will summarise the most important arguments here, a more thorough overview is provided by Croft (2003); Goldberg (1995, p. 9–23) and Michaelis (2003).

Goldberg (1995, p. 9–10) first notes that the lexicalist account has to posit implausible verb senses. The (by now famous) example *He sneezed the napkin off the table* (ibid., at 9, example 8) would force lexicalists not only to include the intransitive meaning of *sneeze* in the lexicon, but also a meaning in which 'someone causes something to move by sneezing'. In a constructivist approach, this verb sense would not be necessary since the semantic frame of 'caused motion' is added by the construction.

Goldberg also argues that the lexicalist approach is circular (p. 10–12). To illustrate her point she lists eight different argument structures for the verb *to kick*, which is part of the problem of multiple argument realisation discussed in section 4.1.1 (also see example 4.5). Lexicalists claim that the case-frame of the lexical entry predicts in which argument structures the verb can occur, so a verb which takes two arguments should occur in transitive constructions. However, the distribution of a verb is also taken as the indicator of which argument structure a verb has. Thus *kick* has a binary argument sense in *he kicked the ball*, a ternary sense in *he kicked him the ball*, etc. Goldberg writes that *"it is claimed that* kick *has a particular* n-*ary sense on the basis of the fact that* kick *occurs with* n *complements [and] it is simultaneously argued that* kick *occurs with* n *complements because it has an* n-*argument sense"* (p. 11). A constructivist could escape this circularity by examining which parts of the meaning are contributed by the verb and which are contributed by the construction (however, no unproblematic proposal has been offered yet, see below).

Goldberg next argues that construction grammars are semantically more parsimonious than lexicalist approaches because there is no need for different verb senses (p. 12–13) and points to experimental evidence supporting her claims from the domains of sentence parsing (p. 17–18) and language acquisition (p. 18–21). She then argues that the traditional motivations for lexical rules (as opposed to transformational syntax) can be reinterpreted from a constructivist point of view. For example, lexicalists argue for lexical restrictions on the argument structure of a verb (e.g. animate vs inanimate). However, (some of) these restrictions are better put in the constructions. The following examples show that *clear* can sometimes occur in the ditransitive depending on the nature of the arguments. In other words, there can be no lexical rule disallowing this verb in the ditransitive construction, but the construction can put restrictions on which arguments can fill its argument slots:

(4.21) Joe cleared Sam a place on the floor.

(4.22) *Joe cleared Sam the floor.
(Langacker, 1991b, quoted from Goldberg, 1995, p. 22, example 32)

Finally, constructivist approaches lead to a new understanding of how language users can ever come to coherent semantic interpretations for creative language use and coercion. For example, the verb *bark* can normally only occur in intransitive clauses, yet speakers of English can perfectly understand what it means in the sentence *A gruff 'police monk' barks them back to work* (Michaelis, 2003, p. 261, example 6b). Here, *bark* occurs in the caused motion construction, which adds its meaning to the whole.

The constructivist alternative. The previous paragraphs already indicated what the alternative is that is proposed by constructivists: here, grammatically relevant meanings or the semantic decompositional structures of events correspond to 'constructional meanings' (Goldberg, 1995, p. 29). Only in some cases verbs share this meaning with constructions (e.g. *give* is a prototypical verb for the ditransitive construction) and recent studies suggest that children acquire the meaning of constructions based on these frequent and prototypical verbs (Goldberg *et al.*, 2004; Goldberg, 2006).

Instead of projecting a verb's argument structure onto syntax through lexical rules or linking rules, constructivists need to explain how verbal meaning integrates with constructional meanings. Goldberg (1995, p. 43) makes a distinction between event-specific 'participant roles' (e.g. a *giver*, a *given*, and a *givee*) and 'semantic roles' (argument roles in her terminology), which are slots in argument structure constructions. Participant roles can be seen as more specific instances of semantic roles.

The lexical entry of a verb specifies which of its participant roles are 'lexically profiled' (i.e. obligatory, not to be confused with profiling in Croft's causal-aspectual analysis, see section 4.1.2). For example, the verb *rob* profiles the robber and the victim of the robbery, but not the robbed goods. The verb *steal* profiles the stealer and the stolen goods, but not the victim:

(4.23) Jesse robbed the rich (of all their money).
*Jesse robbed a million dollars (from the rich).

(4.24) Jesse stole money (from the rich).
*Jesse stole the rich (of money).
(Goldberg, 1995, p. 45, examples 28–29).

Likewise, argument structure constructions specify which of its semantic roles are profiled. For example the English ditransitive construction profiles an agent, a recipient, and a patient. If a verb is member of a class which is conventionally associated with a certain construction, its participant roles may be 'fused' with the semantic roles of the construction (ibid., at 50). This is illustrated in Figure 4.2, which features the representations offered by Goldberg. The top of the figure shows the fusion of the English ditransitive with the verb *hand* as in *I handed him a letter*. This verb has three profiled participant roles (printed boldfaced), which fuse with the three profiled semantic roles of the construction. The bottom of the figure shows the fusion of the ditransitive with the verb *mail*. This verb only has two profiled participants: the mailer and the mailed (e.g. *he mailed a letter*). The participant role 'mailee' is optional and is hence not printed in boldface here. The ditransitive construction however obliges the 'mailee' to be expressed because it fuses with the profiled semantic role of recipient. If a verb

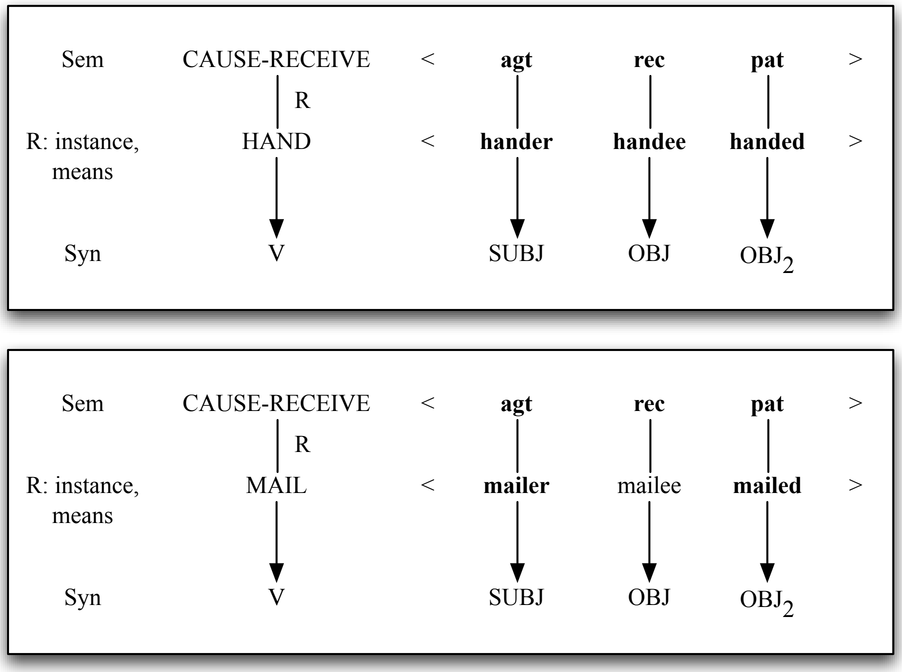


Figure 4.2: These two boxes represent the composite fused structure of the English ditransitive + *hand* (top) and of the English ditransitive + *mail* (bottom). The three profiled participant roles of *hand* can be 'fused' in a one-to-one correspondence with the profiled semantic roles of the ditransitive. The verb *mail*, on the other hand, only profiles the mailer and the mailed (the mailee is therefore not boldfaced). The ditransitive, however, obliges this participant role to be expressed as well (Goldberg, 1995, p. 51 and 53, figures 2.5 and 2.8).

does not have a participant role which can be fused, the construction may add the role itself. For example, *sneeze* only profiles a sneezer but has no other participants. In the caused motion construction, however, the construction adds the semantic roles of goal and theme, which yields utterances such as *he sneezed the napkin off the table*.

Given the observation that constructions carry their own meanings, it may be expected that they are polysemous in nature just like lexical items and morphemes. Goldberg notes that for example the English ditransitive indeed occurs with different related senses, which all have to do with some kind of caused possession (p. 38). She distinguishes the following six meanings (see Croft, 2003; Davidse, 1996, for more senses of the ditransitive):

(4.25) I gave him a book.
→ Central sense: agent successfully causes recipient to receive patient
→ *give, hand, bring, take, serve, ...*

(4.26) I owe him some money.
→ Conditions of satisfaction imply that agent causes recipient to receive patient
→ *guarantee, promise, owe, ...*

(4.27) They refused us access.
→ Agent causes recipient not to receive patient
→ *refuse, deny, ...*

(4.28) He left you a letter.
→ Agent acts to cause recipient to receive patient at some future point in time
→ *leave, bequeath, allocate, grant, reserve, ...*

(4.29) He allowed him a cigarette.
→ Agent enables recipient to receive patient
→ *allow, permit, ...*

(4.30) She baked him a cake.
→ Agent intends to cause recipient to receive patient
→ *bake, make, build, get, grab, win, ...*
(adapted from Goldberg, 1995, p. 38, figure 2.2)

Goldberg presents many compelling arguments for the constructivist approach, but apart from two vague and circular principles (p. 50), she never specifies or operationalises how the 'fusion' process exactly works. I will assess her proposal and operationalise a modified version of it in Chapter 5. In the following paragraphs of this section I will describe another problem with Goldberg's approach to argument structure, which has to do with the debate between lexicalists and constructivists.

A false dichotomy? The constructivist approach allows us to reexamine many of the proposals made in the past. For example, research on lexical decomposition (see section 4.1.2) clearly shows how aspectual and causal considerations play a part in argument linking. However, if constructions have their own semantic representation as argued by construction grammarians, aspectual interpretation of a verb should not be based on aspect in the lexical entry of the verb, but rather on the various tense-aspect constructions of a particular language. For example, the aforementioned sentences *the window shattered* and *the window is shattered* both feature the same verb, but different aspects of the situation are profiled by the English tense-aspect constructions. This would then be another example of how constructions contribute to the meaning of an utterance.

However, not all verbs act in the same way in (what seeminingly are) the same tense-aspect constructions. Bill Croft, one of the greatest proponents of construction grammar himself, warns that constructivists should not throw out the baby with the bathwater when they oppose their theories to lexicalist approaches:

> The lexical rule–construction debate revolves around a false dichotomy. Although constructions exist, as cognitive linguists have argued, the relations between constructions (and their meanings) and their component words (and their meanings) is not as neat or as clear as the lexical rule–construction debate implies. Language once again successfully resists the attempt of linguists to make it neat and clean. (Croft, 2003, p. 50)

Croft illustrates his claim by investigating Goldberg's analysis of the English ditransitive construction (Goldberg, 1995, chapter 6). As shown in examples 4.25 – 4.30, this construction can have several senses, which Goldberg attributes to the polysemy of constructions (p. 31–39). Croft (2003, p. 55) however notes that the variation in the meanings of the ditransitive construction is not true polysemy: for example, the verb *give* can only occur in the sense of 'X causes Y to receive Z' and never in the reading of intended transfer: **I gave him a book, but he didn't accept it.* So each class of verbs only occurs in one sense of the ditransitive.

Croft therefore argues that instead of having one ditransitive construction containing several senses, there must be *verb-class-specific constructions* (p. 58) which each have their own entry in the linguistic inventory. Since the different senses and the syntactic structures are so alike, speakers are still capable of detecting some family resemblance between them.

But verb-class-specific constructions are not enough. First of all, there is evidence that not every verb of permission and refusal (as in examples 4.27 and 4.29) can occur in the ditransitive:

(4.31) Sally permitted/allowed/*let/*enabled Bob a kiss.

(4.32) Sally refused/denied/*prevented/*disallowed/*forbade him a kiss.
(Goldberg, 1995, p. 130, examples 26–27)

These examples mean that these classes are even more fine-graind than proposed by Goldberg (1995). Goldberg also offers some exceptional cases of the verbs *forgive* and *envy* in the ditransitive (p. 132) which need to be accounted for. She argues that these exceptional cases are remnants of old meanings of these verbs: *forgive* used to mean 'to give or to grant' (OED), and *envy* used to mean 'to give grudgingly' (OED). Both verbs thus used to fit in the caused possession frame in earlier times, but their ditransitive uses are now gradually becoming out of fashion:

(4.33) He forgave her her sins.

(4.34) He envied the prince his fortune.
(Goldberg, 1995, p. 132, examples 34–35)

Croft (2003, p. 59) however notes that if these verbs have shifted in meaning, they should no longer be able to occur in the ditransitive. He takes this as evidence that there are – at least for these two verbs – *verb-specific constructions*, which are also independently represented in the linguistic inventory. These verb-specific constructions are instances with the specific verb already specified within it. Given this range from lexical entries and verb-specific constructions to verb-class-specific constructions and (possibly) highly schematic constructions, the dichotomy between the lexical rule approach and constructions is false. Verb-specific constructions are in fact very similar to positing different lexical entries which are tied to argument structure through lexical rules. Croft argues that there are probably many other verbs which have their own, autonomous verb-specific constructions represented in the inventory, which can then drift away from the more schematic constructions in the same way as there is 'lexical split' in languages.

Croft's notion of a 'verbal profile' (see section 4.1.2) should thus be interpreted as this complex interplay between the lexical predicate and the constructions it occurs in. For a combination of a lexical entry and the ditransitive construction, at least three semantic components must be taken into account: (1) the event-type or 'core meaning' of the verb itself, (2) the meaning of transfer of possession that is associated with the verb when it occurs in the ditransitive, and (3) a 'modulation' of this meaning: whether it was actual transfer, intended transfer, etc. (p. 62). As I will argue in Chapter 8, we can in fact do without this third component.

4.1.4 Conclusions

In this first section I gave a brief but critical overview of some approaches of argument realisation and hinted at the complexity involved in argument linking patterns. There are some important conclusions that can be drawn from the discussion:

- A rich representation of event structure is needed to understand the various argument structure patterns that occur in language;
- Semantic roles can only be understood (a) by looking at the participants in the event structure in relation to each other, and (b) by taking the event profile into account (aspect, tense, etc.);
- Argument realisation cannot be specified in one part of the linguistic inventory: instead of positing argument structure information in either the lexicon *or* the grammar, every item in the linguistic inventory can contribute to the meaning;
- Close studies of verbal behaviour and grammaticalization processes suggest that there are not only schematic constructions, but also verb-class-specific and even verb-specific constructions.

In Chapter 5 I will propose a formalisation of how argument realisation may be achieved in Fluid Construction Grammar. This representation first of all forms the basis of the experiments which are described in Part II, but it is also part of an alternative grammar architecture closely related to other construction grammar proposals such as Croft (2001), Goldberg (1995) and Kay & Fillmore (1999).

4.2 A case study on the formation of case markers

It is clear from the myriad of existing theories that argument realisation remains a largely unresolved issue in linguistics. In this dissertation, I will try to obtain new insights into the problem through a detailed study on the formation of a grammar for marking argument structure by doing experiments with a population of autonomous artificial agents. More specifically, I will investigate how these agents can develop a case marking system, comparable to the inflectional category-systems found in languages such as German, Latin, Turkish, and many other languages.

Before diving into the experiments themselves (described in Part II), it is useful to dissect the question of the emergence of case markers into a couple of smaller problems which then can be used as guidance for the experiments and as reference ground for their analysis. First of all, we need a clear hypothesis of *why* every attested language has developed some grammatical way to express argument structure. This hypothesis then serves to decide on the starting point of the experiment. We can then ask ourselves what the subsequent stages of increasing complexity are in the development of case markers that lead to natural language-like functionality. By carefully studying each developmental stage at a time, we can measure every newly added aspect and show its influence on the experimental results.

4.2.1 The functions of case markers

Computational and mathematical studies on the origins of grammar have so far mainly focused on how syntax arises to make the linguistic inventory more optimal in terms of inventory size and learnability. This has especially been true for models of genetic evolution and Iterated Learning Models (see section 3.1). Nowak & Krakauer (1999) report experiments on 'language as an adaptation' in which grammar only becomes necessary when the speakers need to communicate about a large set of events. A somewhat different yet similar approach is found in most experiments in Iterated Learning Models (Smith *et al.*, 2003), which focus on the 'learnability' of language. In cross-generational Iterated Learning Models, languages have to be transmitted from the adult speaker to the child learner. Since language can potentially express an infinite number of sentences, the learner can never observe all possible utterances during a lifetime so there is a 'learning bottleneck'. This bottleneck pressures languages to become 'compositional' rather than 'holistic'. This time, however, compositionality does not become an innate feature of the agents but remains a property of cultural transmission.

It is clear that syntax has an optimisation effect and that the pressure of a 'learning bottleneck' is indeed present in the emergence of language. However, for the reasons given in the previous chapter, these two pressures cannot be the primary triggers of 'true grammar' (at least as it was defined in Chapter 2). As an alternative, I hypothesise that grammar arises in order to (a) reduce the semantic complexity of interpretation for the hearer, and (b) limit the number of possible interpretations (both in order to increase communicative success). In order to apply this hypothesis to case marking systems, we need to have a clear idea of what the functions are of case markers in natural languages.

First of all, case markers can be used as a device for marking what the role of a participant is in a specific event. Consider the following example from Turkish:

(4.35)	*Mehmet*	*adam-a*	*elma-lar-ı*	*ver-di*
	Mehmet.NOM	man-DAT	apple-PL-ACC	give-PAST-3SG

'Mehmet gave the apples to the man.'
(Blake, 1994, p. 1, example 1)

The case markers make it clear to the hearer that Mehmet was the giver of the apples, not the recipient. Marking the roles of participants in an event has some serious advantages: it allows the hearer to interpret the utterance correctly without actually observing the scene (i.e. 'displacement' becomes possible) because it rules out ambiguities, and it reduces the semantic complexity of parsing because the hearer does not have to find out for herself who was the giver and who was the receiver. So one of the basic functions of case marking is **to indicate event structure**.

Secondly, case markers are also used **for packaging information structure**. In example 4.35, the suffix *-ı* not only marks the accusative case, but it also indicates that the apples are 'specific' (as opposed to undetermined). Many languages also exploit case markers for marking the topic of an utterance and the focus, for marking given information versus new information, etc. Information packaging also means the possibility of 'perspectivising' the utterance (e.g. passives can be used to express an event from the viewpoint of the undergoer).

Finally, case marking can be used for indicating various other grammatical distinctions such as number and gender, as is the case in German:

(4.36)	*ein*	*klein-er*	*Man*
	a	little-M.SG	man

'a little man'

(4.37)	*drei*	*klein-e*	*Frau-en*
	three	little-PL	woman-F.PL

'three little women'

Figure 4.3: In the first stage, we only start with lexical items. We assume no semantic roles specified in the lexical entry of a verb, but its own specific participants (for example a 'mover'). The meaning of the verb itself can be decomposed into its event structure.

Languages vary as to what or which grammatical device(s) they use for these functions. For example, English uses word order (and intonation) for marking argument and information structure. Other strategies are verb marking, tone, intonation, adverbs and prepositions. Even though these strategies may be used for similar functions, each of them may have consequences for argument realisation. For example, word order may pressure languages to further grammaticalize their subjects and objects into information structure markers (see sections 4.2.4 and 4.2.5), whereas case marking makes semantic distinctions possible even in subject marking (also see the active-stative languages in section 4.2.4).

In this thesis, I will focus on the formation of a grammar for marking event structure without the influence of other functional pressures such as information structure or other grammatical distinctions. Because semantic distinctions are 'easier' to indicate using case markers than using e.g. word order, I will focus on the formation of a case-like grammatical system for marking argument structure (which then applies to a lesser extent to other grammatical devices as well). I do not claim that marking event or argument structure is the primary function for all case markers (see e.g. the topic marker in example 4.38), but that many existing case markers emerged for this reason. The following sections should therefore be interpreted as the developmental pathway for such markers only.

4.2.2 Stage I: no marking

Given the assumptions explained in Chapter 2 (a) that all grammatical categories are language-specific, and (b) that they are constructed socially through general cognitive mechanisms, the first stage starts with no grammar at all. No grammar also means no semantic roles, such as 'agent' or 'patient', which are considered to be universal or universally derived in many linguistic theories (e.g. Pinker, 1989; Van Valin, 2004) and which are predefined in the meaning of some of the computational and mathematical models I discussed earlier (e.g. Kirby, 2001; Nowak & Krakauer, 1999).

In the experiments described in Part II, it is useful though to start at least from a point where the agents already have a lexicon. This proposal seems to be valid given the many attested cases of grammaticalization which show that new grammatical forms develop from lexical items and later may become part of a construction and develop even further. A second justification for this choice is that there are already many experiments within the same methodology which successfully show how lexicons can be formed in a population of grounded robotic agents (Steels, 1996a,c, 1997e, also see Chapter 3). In this way we can better isolate the features that are hypothesised to be formed in the experiments. Of course, once the dynamics of these experiments are better understood, a series of integrated experiments must be carried out to confirm the results.

But are there attested cases of languages in which the relations between an event and its participants are not marked? One language, Lisu, has been reported which generally does not mark who's doing what in an event. Li & Thompson (1976, quoted from Palmer, 1994, p. 23) give the following two examples:

(4.38) *làma* *nya* *ánà* *khù-a*
tigers TOP dog bite-DECL
'Tigers bite dogs' / 'Dogs bite tigers'

(4.39) *ánà* *xə* *làma* *khù-a*
dog NEW TOP tigers bite-DECL
'Tigers bite dogs' / 'Dogs bite tigers'

In both sentences, there is only a topic marker ('known topic' versus 'new topic'), but the correct reading depends on the context in which the sentence was uttered. Palmer writes that in some constructions, the speakers of Lisu *do* mark the roles played by the arguments in a sentence if needed. He concludes that Lisu *"is sparing in its marking of grammatical roles, though it has a device for doing so when it is really important"* (p. 24). He also notes that even languages such as English do not always mark the distinction between agent and patient. For example, the famous English phrase *the shooting of the hunters* does not indicate whether the hunters did the shooting or whether they got shot themselves. Further evidence is provided by Gil (2008) who reports on languages such as Riau Indonesian which only have very few constructions that mark event structure.

These observations go against the hypothesis that semantic roles are innate linguistic categories which are linked or projected onto syntactic structures. A usage-based approach, however, has no problems in explaining these phenomena and – from the viewpoint of the emergence of grammar – would even predict that all argument constructions start from this stage before additional grammar is added.

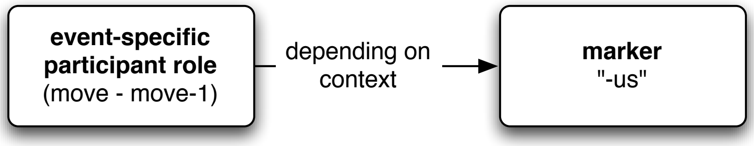


Figure 4.4: In the second stage, a specific marker arises in order to solve a communicative problem. In natural languages, this marker is often a lexical item which is recruited for a new use. In the experiments, this stage involves the invention of a new marker.

4.2.3 Stage II: specific marking

The problem-solving approach to the emergence of language (section 2.1.3) predicts that changes are made in local interactions as a response to specific communicative needs. The second stage would thus be that verb-specific markers are introduced which then may or may not be picked up by others and get propagated in the speech community. In this section I will argue that attested examples in grammaticalization theory and recent empirical evidence provided by Croft (2008) suggest that stage II is indeed a valid step to take.

Construction-specific function. Studies in grammaticalization confirm that innovations occur in specific linguistic contexts: attested examples of the emergence of modern case markers show that they start out in very restricted use. Blake (1994, chapter 6) gives examples of how verbs, nouns and even adverbial particles can develop into case markers. The most fruitful source is probably verbal. Blake writes that a predicate like COME is a two-place predicate implying a 'comer' and a 'destination'. A predicate like LEAVE mirrors this implication by having a 'leaver' and a 'source'. A predicate like FLY, however, only implies a 'flier'. If languages then wish to express utterances such as *he flew to/from Tokyo*, pairs of predicates can be used. He gives the following examples from Thai:

(4.40) *thân cà bin maa krungthêep*
he will fly come Bangkok
'He will fly to Bangkok.'

(4.41) *thân cà bin càak krungthêep*
he will fly leave Bangkok
'He will fly from Bangkok.'
(Blake, 1994, p. 163–164)

Such languages are known as 'serial verb language languages'. The second verb is (usually) non-finite and cannot be marked for tense, aspect or mood independently of the first verb, and it takes no expressed subject or implies the same subject as that of the first verb. Blake writes that functionally speaking, these second verbs are equivalent to prepositions. Serial verb language constructions are in fact very frequent and their development into adpositions and case markers has been widely attested, especially in the languages of West Africa, New Guinea, Southeast Asia and Oceania (ibid., at 163; also see Givón, 1997, section 7, for more on serial verb language constructions and similar examples).

Cross-linguistic studies have shown recurrent patterns in the type of predicates that are recruited in serial verb language constructions and which develop into grammatical markers. These are usually ‘light verbs’ in the sense that they have a very general meaning so they have a broad application and frequent use. Examples are:

(4.42) Examples of adpositions which developed from verbs:

- object
  Yatye *awá* ‘take’, Yoruba *gbà* ‘get’, Mandarin *bǎ* ‘take hold of’
- dative/benefactive
  Yoruba *fún* ‘give’, Mandarin *gěi* ‘give’, Ewe *ná* ‘give’, Thai *hâi* ‘give’
- allative
  Nupe *lō* ‘go’, Mandarin *dào* ‘arrive’, Ewe *ḍé* ‘reach’
- comitative
  Tamil *koṇṭu* (< *kol ‘take’), Mandarin *gēn* ‘follow’
- instrumental
  Mandarin *yòng* ‘use’, *ná* ‘take’, Ewe *tsɔ́* ‘take’
- locative
  Mandarin *zài* ‘be at’, Ewe *lé* ‘be at’
- ablative
  Mandarin *cóng* ‘follow’, Ewe *tśo* ‘come from’
- perlative
  Ewe *tó* ‘pass by’
  (Blake, 1994, p. 166, example 5)

Local interactions. The other prediction, that case markers not only emerge in a specific linguistic context but also in a specific communicative interaction, is much harder to verify. This is due to the fact that linguists most of the time can only notice grammaticalization once it has already occurred, so they usually have to work with ‘symptomatic evidence’:

> [Symptomatic evidence is] evidence from grammatical constraints that are hypothesized to represent conventionalized remnants of ‘pure’ functional constraints. One reason for this limitation is that grammars of natural

> languages generally describe only the conventionalized constraints and not the 'living' functional constraints, which can be uncovered only by extensive text analysis and informant work and so cannot be researched on a large scale. (Croft, 1991, p. 34–35)

Obtaining such empirical evidence is thus hard, but as Croft suggests, not entirely impossible. In fact, Croft himself gathered interesting data in a recent study (Croft, 2008) in which he performs a close-reading of the 'Pear Stories' transcripts (Chafe, 1980). In the original experiment conducted by Chafe, five speakers with similar social backgrounds observed a movie at the same time. Afterwards, each of them had to describe the movie to the experimenter who had not seen the movie him/herself. The experimental set-up was designed in such a way as to reach maximally similar communicative situations. The transcripts thus provide good materials for studying variation in similar interactions by similar people.

Recall that in the utterance-based selectionist view on language (see section 2.3.3), variation in a speech population is the default situation. The model thus predicts that there should also be plenty of variation in the transcripts of the Pear Stories, which may be possible sources of language change. Croft (2008) shows that this is in fact the case and gives examples for various grammatical categories, among which argument marking. Consider the following pair of examples (5 and 7 in ibid.):

(4.43) 2,13 [.75] And a man **comes** along *with a goat*,

(4.44) 19,25 [.45] and a man **comes by leading a goat**.

The speaker of example 4.43 expressed the event using the single verb *come* and a comitative prepositional *with* for the goat. The speaker of example 4.44 verbalised the same event using a second verb *lead* in a non-finite form. The utterance of example 4.44 is remarkably similar to the serial verb language constructions discussed earlier in this section. Here too the second verb is (mostly) non-finite and it has no overt subject. Other speakers used *lead* in finite form, and yet others used either the verb *drag* or *pull* as a non-finite.

Croft then shows examples of similar grammaticalization pathways (from 'take' or 'lead'/'follow' to comitatives) which have been attested in other languages (also see example 4.42). Croft also reports variations in the data that may be the source of grammaticalizations from 'ride' to means of conveyance, from 'give' to a causative or dative, and from 'take' to instrument.

The study thus shows how events can be verbalised in multiple ways, even by one speaker, which causes a pool of variation in the speech community, which then may be the cause of changes in the language. Croft also argues that variation is so pervasive that it is identical to phonetic variation in phonology. He therefore dismisses the hypothesis that speakers only occasionally innovate to be more expressive or to

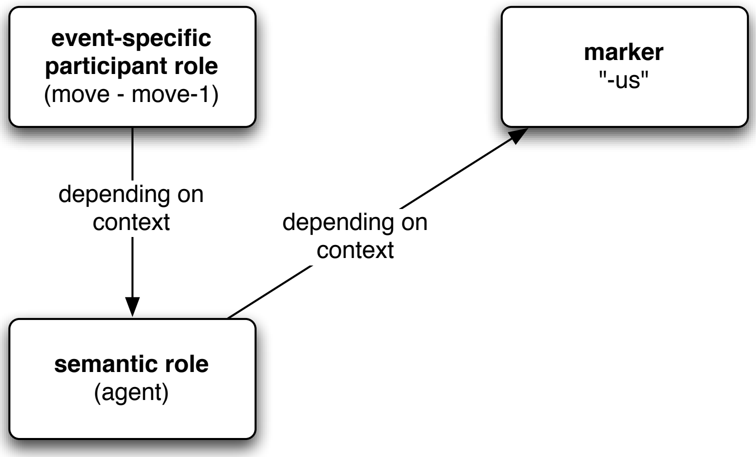


Figure 4.5: Specific markers may get extended through analogy in stage III. They now start to act as semantic roles.

avoid misunderstanding (without of course dismissing them as possible pressures for innovation).

Scaffolding lexical recruitment. The recruitment and evolution of a lexical item into an adposition and eventually a case marker is a long and complex process. For reasons that I will explain in Chapter 6, this crucial step in grammaticalization is 'scaffolded' in the experiments. Rather than reusing an existing lexical item in a more grammatical way, the artificial agents will be capable of inventing a new form which already acts as some kind of adposition or verb-specific case marker. The experiments thus simplify stage II in order to focus first on the function of such markers and how they can be propagated in a speech community. It is needless to say, however, that this part of the grammaticalization chain remains on the research agenda for future experiments and first steps towards grammaticalization of existing lexical elements have already been taken by Wellens *et al.* (2008).

4.2.4 Stage III: semantic roles

If a specific marker invented or recruited in stage II proves to be a successful one, it may extend its use to cover also the arguments of other events. To come back to example 4.40, this would mean that *maa* 'come' evolves from the marker of the destination of a flight to a general allative role (i.e. the destination of motion events). Extension of the use of a marker would in this case be *semantically motivated* and can occur by analogical reasoning over events. This is also the strategy that will be employed in the experiments: once the agents have invented a verb-specific marker, they will be capable of reusing this marker in new situations by analogy.

Since case markers in natural languages usually carry more than one function, it is hard to say what the semantic roles underlying a syntactic pattern are. There are however a couple of diagnostics which can be used, such as looking for 'agnating structures' (Gleason, 1965). Agnation illustrates a structural relationship between two grammatical constructions which have the same (major) lexical items, but different syntactic structures. If the alternation between these two structures is recurrent for groups of constructions, then this can be seen as a pattern in the language. An example of agnating structures is the alternation between the English ditransitive (as in *I gave him a book*) and its prepositional counterpart (as in *I gave the book to him*).

Differences in the semantic categorisation of the verb can come to surface if small but regular variations show up between these agnating structures. Compare the groups of agnating structures in the following examples:

(4.45)	I gave him a book.	I gave the book to him.
	I sent him a letter.	I sent the letter to him.
	I handed him the document.	I handed the document to him.
(4.46)	I baked him a cake.	I baked a cake for you.
	I fixed you a sandwich.	I fixed a sandwich for you.
	I bought him a present.	I bought a present for him.

Even though both groups of verbs can all occur in the ditransitive construction, they select a different preposition in the agnating prepositional construction. The choice for either *to* or *for* is semantically motivated: the verbs listed in example 4.45 entail an *actual* transfer of the direct object, whereas the verbs in example 4.46 indicate that there is an *intended* transfer. This is confirmed in the fact that sentences such as *?I gave him a book, but he refused it* feel awkward, whereas sentences such as *I baked him a cake, but he refused it* are perfectly acceptable. So it seems that both groups of verbs belong to different subclasses.

The extension of specific markers to semantic roles is useful for communication in several ways. First of all, semantic roles increase the potential for generalisation in a language: by extending the functionality of a semantic role, there is a higher chance that it will be reused for categorising new and similar participant roles as well. Instead of having to negotiate a different marker for each new instance, speakers of a language can thus make a semantically motivated innovation which increases the chance that the hearer will understand the speaker's intention. Second, and related to the first point, semantic roles increase the expressiveness of a language because they allow speakers to profile different aspects of the same event. For example, semantic roles can focus on the relation between an agent and a patient as in *he broke the window*, but also profile exclusively the resulting state of one of the participants as in *the window was broken*.

Finally, semantic roles can significantly reduce the inventory size by grouping together larger classes of verbs.

The model proposed in this thesis predicts that the conventionalisation of a mapping between verb-specific arguments and semantic roles (even though semantically motivated) is neither a determined nor a straightforward one. The choice depends on the conventions that are already present in the language, on a speaker's previous experience, frequency, etc. Moreover, the model predicts that there will be several varieties in the population which compete with each other for becoming the dominant semantic role of a particular argument. Thus we can expect that languages come up with very divergent classifications. For example, Italian features dative subjects with verbs such as *like*, whereas German uses its 'normal' nominative case:

(4.47)	*Ich*	*mage*	*Kirsch-en.*
	I.NOM	like-1SG.PRES	cherry-ACC.PL.F.

'I like cherries.'

(4.48)	*Gli*	*piacciono*	*le*	*ciliege*
	they.DAT	like	DET	cherries

'They like cherries.'

(Italian example adapted from Palmer, 1994, p. 27)

Linking similar events can also be reversed from language to language. For example, when French speakers want to say *I miss you*, they literally say something like 'you are missing to me':

(4.49)	*Tu*	*me*	*manque-s.*
	you	me	miss-2SG.PRES.

'I miss you.'

Also a conceptual domain such as space, which has always been considered to be basic to human experience (and hard-wired in the brain), is cut up into a wide variety of categories by the languages of the world (even among closely related languages, Levinson, 2003; Levinson & Wilkins, 2006; Loetzsch *et al.*, 2008). Based on English prepositions and the assumption that the domain of space cannot yield too much cross-linguistic variation, various predictions have been made that spatial categories have very abstract meanings. Perhaps the best known example is the prediction made by Landau & Jackendoff (1993) who claimed that no language should feature a spatial category such as the hypothetical *sprough* with a specific meaning such as 'through a cigar-shaped object'. This prediction has been refuted by the suffix *-vara* from the Californian language Karuk which has exactly this meaning (Levinson, 2003, p. 63). Moreover, Karuk is not a unique case, as can be seen in the following example from South-Eastern Pomo, which features a spatial marker for tunnel-like shapes:

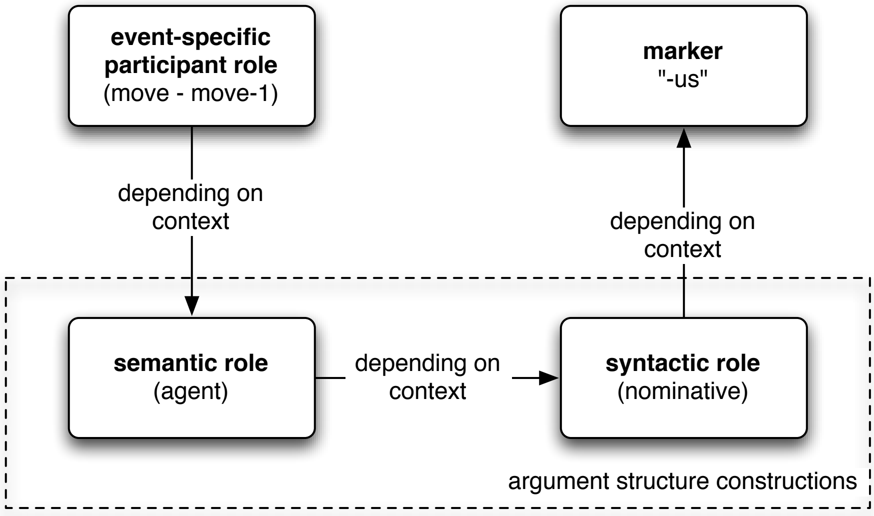


Figure 4.6: In the next step, even more abstractions occur through syntactic roles. The mapping between semantic roles and grammatical cases is hypothesised to be handled by argument structure constructions which can combine several semantic roles into a larger pattern.

(4.50) *lı́l*
into an enclosed space
-bò *-t*
crawling motion along a surface, into a long and horizontal object PAST
'He crawled into a tunnel.'
(Moshinsky, 1974, glosses added by myself based on information found in the grammar)

So semantic roles seem to be language-specific, which fits our assumption that they have to be constructed and learned. But once they become part of a language, are they an affair of 'take it or leave it', or is it possible that the same verb-specific participant role can be mapped onto multiple semantic roles? The answer seems to be that this mapping too is indirect and dependent on the context. For example, the 'sneezer' in *he sneezes* seems to be a patient, whereas it is (also) a causer in *he sneezed the napkin off the table*. All the above observations are reflected in Figure 4.5 and more evidence is provided in the next chapter.

4.2.5 Stage IV: syntactic roles

In the previous section I wrote that it is not easy to detect which semantic roles are distinguished in a language. This is because there is no direct mapping between a semantic role and its surface form. In section 4.1 I also provided evidence that there may be hundreds or more semantic roles, so stage three cannot be the end of the development of a case marker if we examine the functionality of natural language case markers. As Bill Croft puts it:

> It is obvious that, although semantic thematic ROLES may be verb specific, surface morphosyntactic CASE MARKINGS in natural languages gather together large classes of thematic roles since there are thousands of verbs with thousands of roles but never more than fifty to eighty case markers and at the very most just a dozen or so case markers other than spatiotemporal ones. Thus, surface case marking imposes structure on thematic relations to an even more abstract degree than verb roots impose structure on the human experience of events. (Croft, 1991, p. 158–159)

We now need a diagnostic for finding out whether a case marker is still a semantic role or whether it has already developed into a 'syntactic role'. In this thesis, I will say that a case marker has become a syntactic role once it passes the 'dissociation test', which is defined by Givón as follows:

> To demonstrate, even in the most superficial way, that a case-role is grammatical rather than semantic, one must demonstrate its dissociation from semantic roles. That is, one must show that it admits more than one semantic case-role. (Givón, 1997, p. 2–3)

Again, agnation can be used to demonstrate the dissociation of a case marker from semantic roles. As was demonstrated in examples 4.45 and 4.46, the slot of the indirect object in the English ditransitive construction groups together at least six different semantic roles (also see examples 4.25 – 4.30). These semantic roles seem to be semantically related and have to do with 'caused possession'.

Many surface case markers (especially spatiotemporal ones) group together semantically related roles. For example, the Latin dative takes at least nine different roles, but still *"forms a rather homogeneous case: it indicates the pole towards which the action or the process referred to by the predicate is oriented"* (Van Hoecke, 1996, p. 31). Cross-linguistic studies also show that there are strong tendencies in the kind of semantic roles that will be covered by a dative case (Haspelmath, 2003).

This does not mean, however, that there is a single abstract meaning underlying these case markers or that a single definition can be sufficient to cover their application. A more fruitful approach is the aforementioned 'polysemy or usage type approach' (see section 4.1.2). Rather than looking for a single meaning, the functions and uses of case markers are relevant. This not only makes cross-linguistic comparison possible, but it will also form the basis of comparing 'case markers' of artificial languages to those of natural languages (see part II).

But do all case markers cover only semantically related roles? In English, the roles of subject and object can cover virtually any semantic role (a similar observation goes for the nominative and accusative case in many languages):

(4.51) Multiple semantic roles of the grammatical subject:

a. Patient of state:
   She is tall.
b. Patient of change:
   She is falling asleep.
c. Dative:
   She is dreaming.
d. Agent:
   She is writing a letter.

(4.52) Multiple semantic roles of the grammatical object:

a. Patient of state:
   He saw her.
b. Patient of change:
   He pushed her.
c. Ablative:
   He approached her.
d. Allative:
   He left her.
e. Ingressive:
   He entered the house.
f. Dative:
   He gave her a book.
g. Benefactive:
   He built her a house.

(Givón, 1997, p. 3, examples 1 and 2)

So the English subject and object position group together unrelated semantic roles. Many theories therefore make a distinction between 'grammatical' or 'core' cases on the one hand, and 'semantic', 'concrete' or 'peripheral' cases on the other (Blake, 1994, p. 32–34). This distinction is defined in many ways by various authors, but the main idea is that 'grammatical cases' (typically nominative and accusative; or subject and object) seem to be syntactic markers which are dissociated from any semantics, whereas for semantic cases (typically spatiotemporal ones) some semantic meaning can still be revealed.

The question is whether such a distinction is useful for our purposes, that is, are there good reasons for making a distinction between the evolution of grammatical cases versus the emergence of semantic cases? The answer is no: cross-linguistic and diachronic evidence suggests that there is no clear-cut separation possible between grammatical and semantic cases because there are many gradations in the grammaticalization process of a case marker. For example, Givón (1997, p. 25–26) shows that direct objects typically extend their functionality by taking over semantic roles that used to be covered by other cases such as the dative and associative (see Table 4.2). In other words: cases differ from each other in degrees of grammaticalization and functionality so there is no reason why we should assume that speakers make the distinction between grammatical and semantic cases.

Evidence that the distinction is not relevant to speakers can be found in typological data. As seen in example 4.42, the adposition *awá* 'take' developed into an object-marker in Yatye, whereas similar adpositions developed into comitative or instrumental markers in Mandarin, Ewe and Tamil (also see section 4.2.3). This suggests that at least parts of the developmental pathways of case markers run along parallel lines. Other evidence can be found in the various case alignment strategies observed in natural languages. One type of languages called 'active-stative' languages (e.g. Lakota, Choctaw, Guaraní, Eastern Pomo) feature a subject (mostly of intransitive clauses) that can have various markings depending on the semantics or aspect of the verb. For example, Choctaw has agreement markers for nominative, accusative, dative and benefactive:

(4.53) Nominative:

-li	*ish-/is-*	*il-/i:-*	*hash-/has-*
1SG NOM	2 SG NOM	1 PL NOM	2 PL NOM

(4.54) Accusative:

sa/si-	*chi-*	*pi-*	*hachi-*
1SG ACC	2 SG ACC	1 PL ACC	2 PL ACC

(4.55) Dative:

(s)am-(s)ã-	*chim-/chĩ-*	*pim-/pĩ-*	*hachim-/hachĩ*	*im-/ĩ*
1SG DAT	2 SG DAT	1 PL DAT	2 PL DAT	3 SG/PL DAT

	language	type of object role
		most semantic
a.	Japanese	patient
b.	Hebrew	definite patient
c.	Spanish	dative, human patient
	Provencal	dative, pronoun patient
	Newari	dative, topical patients
d.	Ute	patient, dative, beneficiary
e.	Totzil	patient, dative, beneficiary, possessor of object
f.	English	direct object
	Nez Perce	
	Kinya Rwanda	
		most grammatical

Table 4.2: Grammaticalization is a gradient process, which is illustrated here for direct objects in some languages. This table shows from top to bottom a continuum of languages with a more ‘semantic’ direct object to languages with a more ‘grammatical’ or ‘syntactic’ direct object (table adapted from Givón, 1997, p. 26, table (46)).

(4.56) Benefactive:

(s)ami-	*chimi-*	*pimi-*	*hachimi-*	*imi-*
1SG BEN	2 SG BEN	1 PL BEN	2 PL BEN	3 SG/PL BEN

(Davies, 1986, p. 3, example 1)

The grammaticalization of case markers into such semantically coherent classes is represented in Figure 4.6 which is a grammatical square involving the participant roles, semantic roles, syntactic roles (or grammatical cases), and surface forms for case markers. Each mapping in the square can vary according to the communicative and linguistic circumstances, the relation between a participant role and its morphosyntactic realisation is multilayered and indirect.

The picture presented here however has the danger that case markers are considered in isolation, whereas I argued in the first section of this chapter that they can only be understood in relation to the other parts of the pattern in which they occur. Indeed, the mapping of semantic roles onto syntactic roles (and vice versa) is hypothesised to be taken care of by argument structure constructions. These constructions may be schematic, verb-class-specific or even verb-specific. I will come back to this point in Chapter 7 in which I present experiments investigating the formation of such larger argument structure constructions.

As I will demonstrate in the experiments in Part II of this thesis, the combinations of case markers into patterns or constructions may have important consequences for the lifetime and evolution of the markers that are involved: they are no longer independent items in the language and they therefore have to rely on each other or find a more successful 'partner' for becoming a conventionalised unit in the language. I did not include a separate stage for the formation of combinations of case markers because this may occur at various places in their development and stop or accelerate the transition from one stage to the next.

4.2.6 Stage V: further developments

Stages I–IV described the possible evolutionary pathway of a case marker which resulted in an inflectional category that groups together various semantically related roles. This is also the endpoint of the experiments described in this thesis because they focus only on case marking as a way to express event structure in a grammatical way. There are, however, other functions which may be performed by case markers such as packaging information structure, marking perspectivisation and other grammatical distinctions such as gender and number (as argued in section 4.2.1). Especially information structure seems to be the most important pressure for case markers to extend their coverage and become even more dissociated from their previous meaning (see Table 4.2).

Finally, case markers eventually decline and even whole grammatical systems of case may get lost and replaced by other grammatical devices. This is apparent in the (almost) complete loss of case marking in languages such as English, French and Dutch, which replaced it with word order constraints. Individual case markers may disappear because they are 'merged' with another case such as the merger of the instrumental and locative case in Middle Indo-Aryan (Blake, 1994, p. 176). Whereas merger seems to be relatively common in the life cycle of case systems, case split is quite rare. Merger of case markers means that the case system is reduced unless new members are recruited. With the loss of cases and further developments in the grammaticalization of case markers, the different cases of a language may become insufficiently differentiated from each other. This allows other strategies, such as word order complemented with prepositions in English, to become more popular and eventually the new conventions of a language.

The decline of entrenched and conventionalised grammatical units is also not the focus of the experiments reported in this dissertation. It is needless to say, however, that the loss of grammatical devices to the advantage of others is a very interesting research topic and even crucial to our understanding of emergent grammar. However, we first need to fully understand the development of one grammatical system before we can investigate how various strategies may compete with each other.

4.3 Summary: the grammatical square

In this chapter, I further substantiated the theoretical foundations of this thesis by applying the assumptions of Chapter 2 to grammatical devices for marking argument structure. Since computational experiments on the emergence or formation of such grammatical devices should (a) come up with an adequate representation of argument realisation, and (b) have a clear idea of what the developmental stages of these devices are, I zoomed in on both aspects through a theoretical discussion which can both be seen as specifying the requirements which have to be operationalised by the experiments as well as the reference ground for comparing emergent artificial languages to natural language grammars.

In the first section, I illustrated the problem of argument realisation by reviewing three different approaches: semantic role lists, lexical decomposition and event structure, and constructivism. I pointed to the insights gained by these theories as well as to their shortcomings, and summarised the most important conclusions that every theory of argument realisation has to take into account. I will offer an alternative proposal and formalisation within a construction grammar framework in Chapter 5.

Next, I delineated the scope of the experiments by specifying that they focus on the formation of a case-like system which functions uniquely as a way to express argument structure. Other possible functional pressures such as information structure and perspectivisation are therefore not investigated. I then dissected the evolution of case markers into subsequent developmental stages. The experiments reported in Part II of this dissertation will focus on the first four stages in this development, ranging from an entirely lexical language to a language featuring 'syntactic' cases (in the sense that they cover more than one, but still semantically related roles). I presented the notion of the grammatical square which illustrates the indirect and multilayered relation between meanings and their forms.

Part II

Formalisation and Experiments

Chapter 5

Argument Realisation in Fluid Construction Grammar

In section 4.1.3 I gave a brief overview of the main reasons for adopting a constructivist approach to argument structure but I didn't specify exactly *how* construction grammar integrates lexical and constructional meanings without using lexical rules (or projection or linking rules). The main reason for postponing this matter to this chapter is that the formalisation of 'argument linking' is still a matter of debate in the field. In fact, whereas lexicalist approaches can boast a large number of computational implementations (Combinatorial Categorial Grammar (CCG), Steedman, 2000; Lexical-Functional Grammar (LFG), Bresnan, 1982; Head-driven Phrase Structure Grammar (HPSG), Pollard & Sag, 1994; etc.), computational construction grammars are scarcely out of the egg: the formalisation of Berkeley Construction Grammar (e.g. Kay & Fillmore, 1999) has not actually been implemented (yet); Embodied Construction Grammar (ECG, Bergen & Chang, 2005) can parse utterances, but no successful production mechanism has been reported yet; and the first attempts to marry HPSG with (Berkeley) construction grammar (called Sign-Based Construction Grammar) were still in preparation at the moment this dissertation was written (Fillmore *et al.*, in preparation).

The work in this thesis uses yet another grammar formalism called Fluid Construction Grammar (FCG). FCG has primarily been developed to support research on the formation of grammar through computational simulations and robotic experiments. While these experiments involve the evolution of artificial languages, the linguistic relevance of FCG as a computational implementation of construction grammar has always been a point of attention. This becomes apparent when FCG is applied to non-trivial problems such as the formation of a case grammar in which a working solution is needed for integrating lexical and constructional meanings. In fact, one of the main challenges of the experiments reported in this thesis has been designing the proper representation and operationalisation of how lexical entries can interact with argument structure constructions.

In this chapter I present this operationalisation and apply it to some problems of argument linking found in English. To my knowledge, this solution contains some novel ideas which may be relevant for describing natural languages as well. I will come back to this claim in Chapter 8 where I compare my solution to a recent proposal on argument realisation in Sign-Based Construction Grammar. The main point of this chapter is however neither how FCG handles argument realisation in natural languages, nor is it my purpose to offer the best or most general description of any particular construction. This would require much more research and could be the topic of a monograph in itself. This chapter rather demonstrates how the FCG formalism is used in the experiments reported in the rest of Part II.

This chapter is structured as follows: first I will show that there is a need in construction grammar to formalise and operationalise how lexical entries integrate with constructions. I will do so by reviewing Adele Goldberg's 'fusion' process and illustrate its shortcomings. The rest of the chapter then deals with how this problem can be solved using Fluid Construction Grammar. First, I give a quick overview of how meaning is represented in this thesis and how meanings can be linked to each other using this formalisation. In section 5.3, I will then give a brief overview of how FCG handles linguistic processing. Finally, I will present a few hand-made examples that illustrate the ideas put forward in this thesis and how constructions are used in the experiments.

5.1 Integrating lexical and constructional meanings

Construction grammar breaks with the traditional idea that the lexicon and syntax are two separate modules and proposes that all linguistic knowledge can be represented in a structured taxonomy of constructions, also known as the 'constructicon'. As explained in the previous chapter, this approach forces us to look at argument realisation in a different way: instead of using lexical rules or projection principles, a strategy has to be found that explains how lexical and constructional meanings can integrate with each other to license acceptable utterances.

As for the interaction between verbs and argument structure constructions, Goldberg (1995, p. 50) proposes that the participant roles of the verb have to be 'fused' with the semantic roles of the construction(s), which basically means that the (grammatically relevant) meanings of both have to be compatible with each other. Unfortunately, Goldberg only offers two vague principles for allowing or disallowing fusion (quoted from ibid., at 50):

> 1. *Semantic Coherence Principle:* Only roles which are semantically compatible can be fused. Two roles r_1 and r_2 are semantically compatible if either r_1 can be construed as an instance of r_2, or r_2 can be construed as an instance of r_1. For example, the kicker participant role of the *kick* frame may be fused with the agent role of the ditransitive construction because

> the kicker can be construed as the instance of the agent role. Whether a role can be construed as an instance of another role is determined by general categorization principles.
>
> 2. *The Correspondence Principle:* Each participant role that is lexically profiled and expressed must be fused with a profiled argument role of the construction. [...]

There are serious empirical problems with these principles. For example, languages feature many verbs that roughly denote the same event type but which have different linking patterns. Well-known examples are *Death frightens me* versus *I fear death*, and *This pleases me* versus *I like this*. In each pair, one pattern links the cause of fear or pleasure to the subject position, whereas the other pattern selects the experiencer as subject. The two principles therefore do not capture why the linking of these verbs is consistent through their use and why we never encounter sentences such as **I feared him* meaning something like 'I caused him to be frightened'.

The only way out seems to be convention as Goldberg suggests herself: *"If a verb is a member of a class that is* **conventionally** [my stress, RvT] *associated with a construction, then the participant roles of the verb may be semantically* fused *with argument roles [i.e. semantic roles] of the argument structure construction*" (p. 50). Convention or entrenchment can explain why the acceptability of utterances may differ from person to person. For example, Pinker (2007, p. 75) quotes columnist David Brooks from the New York Times who uses the idiom *kiss it goodbye* in a prepositional alternation:

> You can kiss goodbye, at least for the time being, to some of the features of the recent crises. You can kiss goodbye to the fascinating chess match known as the Middle East peace process... You can also kiss goodbye to the land-for-peace mentality. [original quote from the *New York Times*, "As Israel Goes for Withdrawal, Its Enemies Go Beserk", July 16, 2006]

Pinker himself, however, finds this use of the idiom unacceptable, as he writes: *"I've read these sentences over and over, but they still stick in my craw. So do many other edgy constructions I have heard or jotted down"* (p. 75). He then gives another nice example of how young Macintosh users can *sleep the computer* (p. 75–76) instead of *switching the computer to sleep mode*. The verb *sleep*, then, does not only have an intransitive sense, but it may also occur in a particular transitive idiom and in a caused-motion idiom as in *he slept the night away*. The subject of *sleep* can thus be mapped onto an experiencer, an agent or a causer depending on which construction it occurs in.

Convention can thus help to solve the problem noted by Croft (2003) that a verb or a class of verbs can only occur in one of the 'polysemous' senses of a construction (see the paragraph 'a false dichotomy' in section 4.1.3). But convention is not enough, as can be seen in the following examples:

(5.1) She drove the car to him.

(5.2) She drove the car for him.

(5.3) *She drove him the car.

Examples 5.1 and 5.2 have the same form as the prepositional alternations of two senses of the English ditransitive construction but have nothing to do with caused or intended possession: *to* has a directional sense ('goal' instead of 'recipient') and *for* has a benefactive sense but without reference to possession (she drove the car at his request or in his place, not with the intention of giving it to him). The fact that 5.3 (meaning that she drove the car in order to give it to him) did not conventionalise is thus functionally motivated: it would make the interpretation of the other two examples ambiguous between a 'possession' and a 'motion' sense. In line with our assumptions on the functions and triggers of grammar, this kind of ambiguity is something that speakers want to avoid if possible.

Similarly, the verb *sweep* does not occur in the English ditransitive because it is associated with motion, even though a sentence like **I swept him the dust* is perfectly intelligible. So even though a sweep-event might lend itself to a possessive reading, the speakers of English prefer not to do so (at least not by using the ditransitive). At first sight, this seems to be evidence that only the lexical entry of the verb should be constrained, not some grammatical construction. However, this solution would still struggle with examples such as the following in which a verb *can* occur in a particular structure, but not with just any argument (5.4 – 5.6 provided by Laura Michaelis, pers. comm.):

(5.4) He sweeps the floor.

(5.5) He sweeps the dust off the floor.

(5.6) *He sweeps the dust.

(5.7) *He sweeps the floor out of the room.

Lexicalists would have to posit two lexical entries for *sweep* or an additional derivational lexical rule with each their own specific feature restrictions to account for these examples. Constraints in the construction would thus capture this alternation more elegantly. More important than elegance, however, is the fact that both uses of *sweep* may denote the same event, but only with different profiles. In this chapter I will therefore propose a solution in which the same lexical entry can combine with several constructions yielding different event profiles or different verb senses. This solution not only solves the problem of 'fusion' but also surpasses the dichotomy between lexical rules and constructions: the lexical entry of the verb is 'tightened' through language use but not so much that it dictates how the argument structure should look like. Instead, the lexical entries offer **potential valents** from which constructions can select an **actual valency**. I will illustrate this implementation for examples such as 5.4 – 5.7.

5.2 Representing and linking meanings

Before going into the technical details of the implementation, it is useful to have a quick overview of the analysis that underlies the examples and how meanings are represented and linked to each other. In the experiments, I will follow the formalisation proposed by Steels (2005b, also see De Beule, 2007; Steels *et al.*, 2005; Steels & Wellens, 2006, for more work using this formalisation) and use a logic-based representation for meaning. For example, the utterance *box and ball* may be represented as follows:

(5.8) box(?obj-x) $\wedge$ ball(?obj-y)

As is standard practice in first order predicate calculus, logic variables (always indicated with a question mark in this thesis) are used for referring to objects. The logic variable '?obj-x' thus has to be bound to the object [BOX] and the variable '?obj-y' has to be bound to the object [BALL]. Suppose that we want to say that the box is big and that the ball is blue, then the meaning may be represented as:

(5.9) big(?obj-x) $\wedge$ box(?obj-x) $\wedge$ blue(?obj-y) $\wedge$ ball(?obj-y)

Note that 'big' and 'box' share the same variable '?obj-x' because they both refer to the same object; and that 'blue' and 'ball' also share the same variable '?obj-y' because they both refer to [BALL]. The speaker may then express this meaning as *(the) big box and (the) blue ball*. Now imagine that the hearer is a non-native-speaker of English who has learned several words, but hasn't acquired the grammar yet. In other words, she does not know that English uses word order in an Adjective-Noun Construction to indicate which adjective modifies which noun. If she just uses her limited linguistic knowledge of English (i.e. the lexical entry of each word), she would come up with the following parse:

(5.10) big(?obj-w) $\wedge$ box(?obj-x) $\wedge$ blue(?obj-y) $\wedge$ ball(?obj-z)

In this parsed meaning, there are neither shared variables between *big* and *box*, nor between *blue* and *ball* because lexical meanings (in our approach) do not specify which words go together. So there may be several hypotheses possible: each word may refer to a different object (i.e. in some languages adjectives can be used as heads of a phrase as well as nouns), *blue* may refer to [BALL] but also to [BOX] (as it is possible to put adjectives in French both before and after a noun as in *un grand ballon bleu* 'a big blue ball', lit. 'a big ball blue'), etc. So the hearer has to witness the scene if she wants to disambiguate the possible interpretations of the speaker's utterance, which may even not be possible if there are other big and blue objects available.

One of the primary functions of grammar is therefore marking which meanings should be linked to each other. In the present formalisation, this means that the grammar has to take care of **variable equalities**. Variables are said to be equal if they refer to the same object. In the meaning in example 5.10, there is an equality between '?obj-w' and '?obj-x' because they both refer to [BOX], and there is an equality between '?obj-y' and '?obj-z' because they both refer to [BALL]. These variable equalities can be resolved by the English Adjective-Noun Construction, which leads to the meaning of example 5.9.

Linking events and their participants. As argued in Chapter 4, one of the main functions of a grammar for argument structure (and the only one I will focus on in this thesis) is making clear who's doing what in an event. This can be easily operationalised using the same formalisation of linking meanings through variable equalities. I will illustrate this for examples 5.4 and 5.5 which are two different argument realisation patterns of the verb *sweep*.

If we consider *sweep* to be a verb of surface contact and motion (such as *wipe, rub,* and *scrub*; see Levin & Rappaport Hovav, 1999), then *sweep* can take at least three participant roles: a sweeper, something being swept, and the source from where the motion starts. One could imagine other roles as well, such as the instrument used for sweeping (a broom or a hand) or the destination of the motion, but they are not necessary for this discussion. In a logic-based representation the meaning of *sweep* can be represented as follows:

(5.11) sweep(?event-x) ∧ sweep-1(?event-x, ?object-a) ∧ sweep-2(?event-x, ?object-b) ∧ sweep-3(?event-x, ?object-c)

Note that the meaning does not only contain a logic predicate for the event, but also explicit predicates for the participant roles themselves. An explicit representation of the verb-specific participant roles is in line with the constructivist approach (see section 4.1.3) and allows us to integrate the verb's meaning with constructional meanings more easily (see below). Instead of using the labels *sweeper*, *swept* and *source*, the more neutral labels *sweep-1*, *sweep-2* and *sweep-3* are used. These are in fact arbitrary labels which can be mapped by robotic or software agents onto their sensory experiences. The meanings of the words *jack*, *floor*, and *dust* can be represented as:

(5.12) jack(?object-u)

(5.13) floor(?object-v)

(5.14) dust(?object-w)

These meanings are introduced by the lexical entry of each word, but it is not yet specified how the meanings should be linked to each other. This is again taken care of by the grammar through variable equalities, which is illustrated in Figure 5.1. Here, the variables ?object-u and ?object-a should be made equal because they both refer to [JACK]. The equality between ?object-v and ?object-b should also be identified because they both refer to [FLOOR]. After making the coreferring variables equal, the hearer can parse the following meanings for examples 5.4 and 5.5:

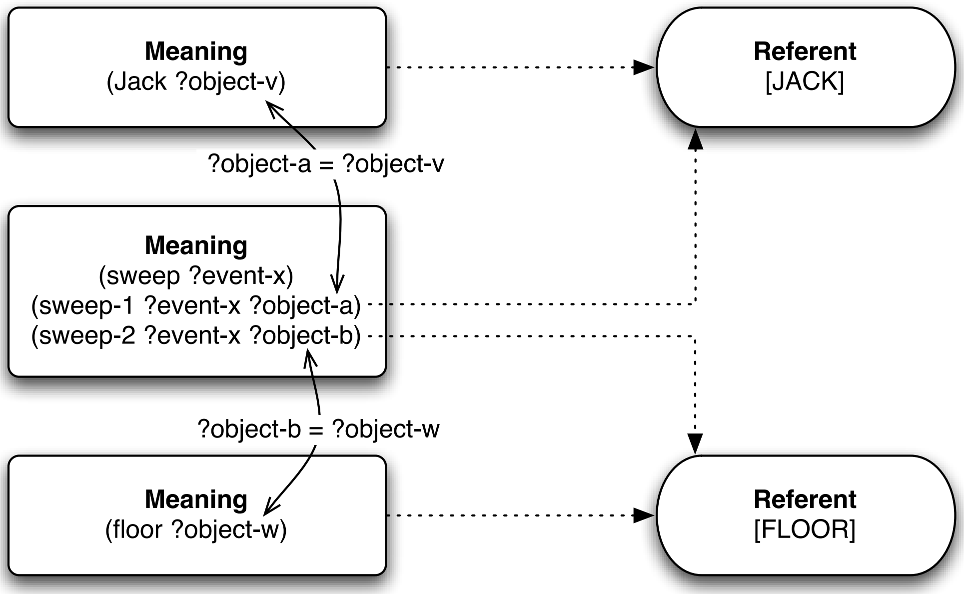


Figure 5.1: One of the primary functions of grammar is to indicate how different meanings should be linked to each other. In the present formalisation, this is represented through variable equalities: the variables '?object-u' and '?object-a' should be made equal because they both refer to [JACK]; and the variables '?object-v' and '?object-b' should be made equal because they both refer to [FLOOR].

(5.15) ∃ ?object-a, ?object-b, ?object-c, ?event-x: jack(?object-a) ∧ floor(?object-b) ∧ sweep(?event-x) ∧ sweep-1(?event-x , ?object-a) ∧ sweep-2(?event-x, ?object-b) ∧ sweep-3(?event-x, ?object-c)

(5.16) ∃ ?object-a, ?object-b, ?object-c, ?event-x: jack(?object-a) ∧ dust(?object-b) ∧ floor(?object-c) ∧ sweep(?event-x) ∧ sweep-1(?event-x, ?object-a) ∧ sweep-2(?event-x, ?object-b) ∧ sweep-3(?event-x, ?object-c)

As will be discussed later, the mapping between these meanings and their forms are organised by constructions through a layer of semantic and syntactic categories. In the mini-grammar presented here, example 5.4 involves a construction which maps the semantic frame 'AGENT-ACTS-ON-SURFACE' onto an SVO-pattern, and example 5.5 involves a construction which maps the semantic frame 'AGENT-CAUSES-PATIENT-TO-MOVE' onto the syntactic pattern SVO-Oblique.

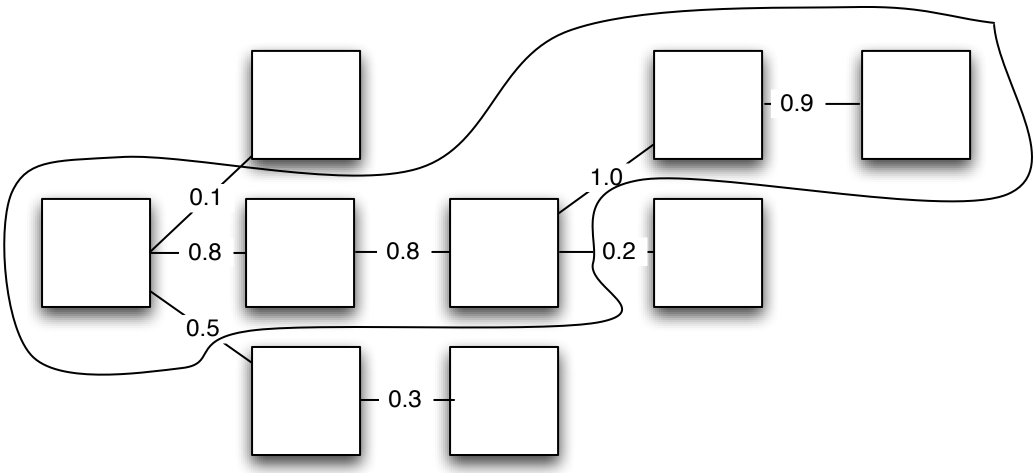


Figure 5.2: During processing, the language user builds up a reaction network by trying out constructions. The pathway with the highest confidence scores is selected for producing or parsing an utterance.

5.3 Production and parsing in Fluid Construction Grammar

Since FCG is interested in grounded language use, a lot of attention goes to the processing of utterances both in production and parsing. As such, FCG can be categorised as a 'procedural language processor' in which processing and grammatical knowledge are tightly integrated. This procedural mode of operation differs from declarative grammar formalisms such as HPSG that want to provide a minimal model of what linguistic entities are possible in a language rather than how they are processed (Pollard & Sag, 1994). In declarative grammar formalisms, grammar and processing are thus separated from each other, although it is possible to compile the grammar into a parser or generator to speed up processing.

During processing in FCG, a speaker or hearer builds up a 'reaction network' or search tree (see Figure 5.2) in which each node represents a stage in the build-up of the semantic and syntactic structure of an utterance. Travelling from one node to the next can be achieved by using a construction. Since there may be several hypotheses given a certain context, each link between the nodes has a 'confidence score' to guide the search. This score is based on (a) the (linguistic) context in which constructions can be applied, and (b) how successful the applied constructions have been in previous communicative situations. The language user will in the end choose the chain with the highest estimated success. FCG is a data-driven and non-deterministic language processor, which means that agents can backtrack in the search tree if necessary.

FCG uses many well-known techniques from computational linguistics such as unification and feature structures to represent linguistic knowledge. In fact, all linguistic knowledge (including constructions) is represented as **coupled feature structures**, which couple a semantic pole to a syntactic pole. All feature structures are organised in units which are used by the basic operators of FCG for retrieving and adding new feature-value pairs. Thus, all linguistic knowledge is represented according to the following pattern:

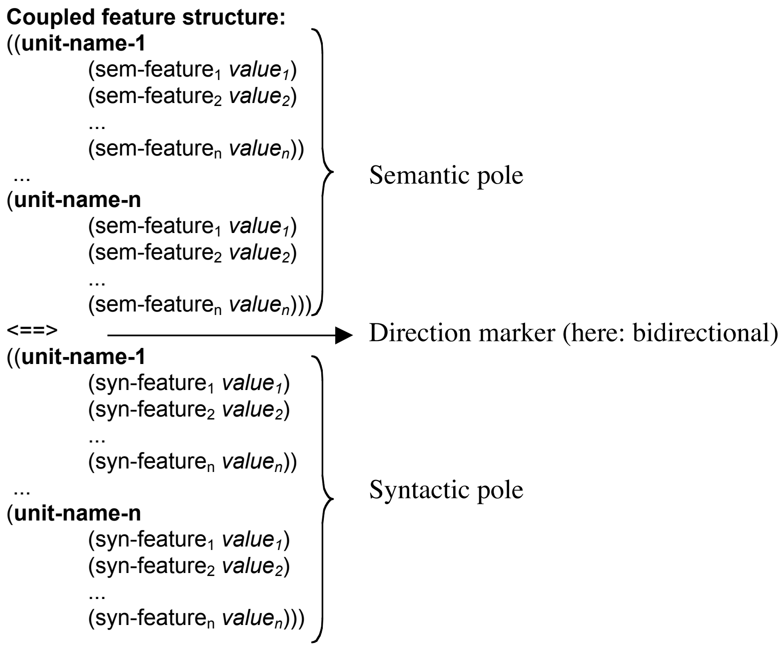


Figure 5.3: Fluid Construction Grammar organises all linguistic knowledge into coupled feature structures.

Language processing. The two basic operations performed in FCG are called ‘unify’ and ‘merge’ (Steels & De Beule, 2006, not to be confused with ‘merge’ in Minimalism). These operators decide whether or not a construction licenses the transition from one node in the reaction network to the next one. ‘Unify’ means that – depending on the direction of processing – the feature structure of one of the poles of a coupled feature structure acts as a set of constraints which have to be compatible with the corresponding pole of the current node in the reaction network. If all the constraints are satisfied, the other pole is ‘merged’ with the corresponding pole in the current node (unless merging fails because of conflicts in both poles). The combination of the unify and merge operations leads to a new coupled feature structure, which is the next node in the reaction network (see Figure 5.4).

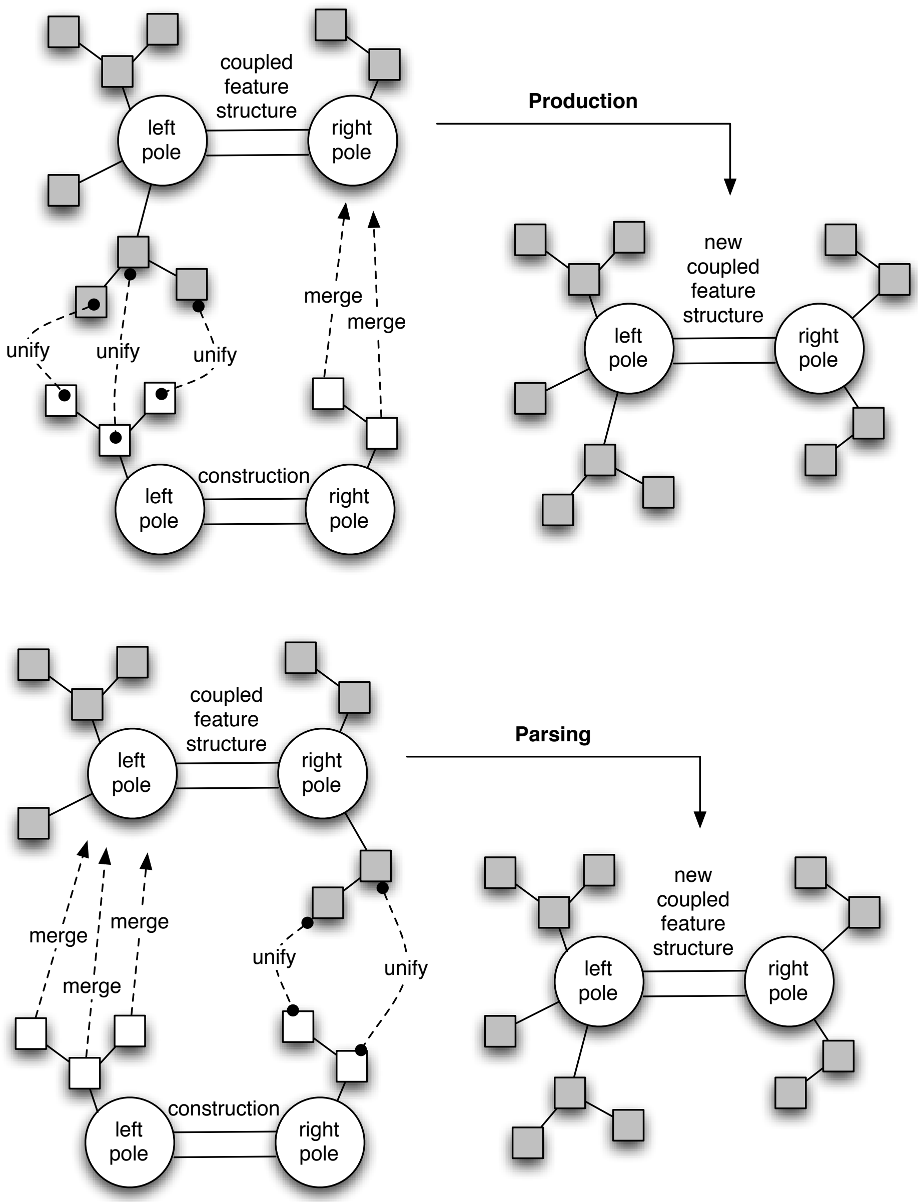


Figure 5.4: During production, the feature structures (squares) in the left pole of a construction (bottom left in top figure) are unified with those of the current coupled feature structure (top left in top figure). If unification is successful, the right pole is merged with the coupled feature structure. This yields a new coupled feature structure, which is a new node in the reaction network (right). During parsing, the same operation is performed but this time the right pole of the construction is unified and the left one is merged with the current node.

Constructions can be applied bidirectionally in FCG, so parsing and production use the same linguistic knowledge but in opposite directions. This is also different from other grammar formalisms. In declarative, non-procedural grammars no directionality is specified but production and parsing routines act as interpreters of the grammatical knowledge. In compiled processors, the grammar is compiled into separate production and parsing modules. Achieving both production and parsing is a non-trivial problem, and some formalisms therefore focus exclusively on either production or parsing (e.g. Embodied Construction Grammar). FCG makes no claim that *all* linguistic knowledge is bidirectional, but uses bidirectional application of constructions for coupling input to output so that agents can act both as speakers and as hearers.

When in production mode, the speaker unifies the left pole of a construction (typically the semantic pole) with that of the current node; and if successful she merges the right pole (typically the syntactic pole) with that of the current node. During parsing, exactly the same linguistic items are used, but this time in the opposite direction: here, the right pole has to unify with the corresponding pole in the current node after which the left pole can be merged with its corresponding pole in the current node.

Structure building. FCG also features structure building operators, which can be used to specify hierarchical relations between units. In these examples, the 'J-operator' is used in the lexical entries and constructions. Roughly speaking, all units marked with a J-operator are ignored during the unification phase, but are added or merged during the merge operation. I will limit my discussion here to the functions of the J-operator that are relevant for the constructions in this chapter. For a more technical specification, see De Beule (2007, chapter 4) and De Beule & Steels (2005). The basic syntax of the J-operator looks as follows:

```
((J ?unit ?parent (?child-1 ... ?child-n))
  (feature-1 value-1)
  ...
  (feature-n value-n))
```

In this syntax, the variable '?unit' has to be bound to some unit in the coupled feature structure. If the unification phase did not yield a binding for this variable yet, a new unit will be created. This unit is then made a sub-unit of the unit that matches with the second argument of the J-operator. In this example this is the unit that is bound to the variable '?parent'. The J-operator can also take an optional third argument, which is a list of the units which have to be made sub-units of the unit that is bound to the first argument of the J-operator (?unit). Next, additional feature-value pairs can be listed which are merged to the structure by the J-operator. I will illustrate the J-operator through an example. Suppose that we only have the top-unit on one of the poles in the coupled feature structure and that there is a construction containing the following J-unit:

```
((J ?new-unit ?top-unit))
```

Since there is no unit yet which could have been bound to the variable ?new-unit during the unification phase, a new unit is created. This unit is then made a sub-unit of the unit that is bound to the second argument of the J-operator so that we get the following structure:

(5.17)

```
  Top-unit
     |
  New-unit
```

Suppose now that another construction applies which contains the following J-unit:

```
((J ?another-new-unit ?top-unit (?new-unit)))
```

Let's assume that the variable ?top-unit was bound by the unifier to 'top-unit' and that the variable '?new-unit' was already bound to 'new-unit' (which depends on the unification of the units that were not marked by the J-operator). The variable '?another-new-unit' does not have a binding yet so a new unit is created which is made a sub-unit of 'top-unit'. This time there is also a third argument of the J-operator. All the units in this list are made sub-units of the newly created unit so we get the following structure:

(5.18)

```
     Top-unit
        |
 Another-new-unit
        |
     New-unit
```

Another use of the J-operator that is adopted in this thesis is the possibility of merging additional feature-value pairs to an existing unit. This can be done as follows:

```
((J ?unit NIL)
  (feature-1 value-1)
  ...
  (feature-n value-n))
```

If the second argument of the J-operator is an empty list (NIL), no structure building operation needs to be performed. In this case, the J-operator will only try to merge the feature-value pairs that are specified with the structure of the unit which is bound to the first argument of the J-operator (?unit). This functionality is needed because it allows the merging of new feature-value pairs on the same pole after successful unification rather than only merging the other pole of the construction to the coupled feature structure. In other words, it allows a construction to add both semantic *and* syntactic feature-value pairs to the coupled feature structure.

FCG does not represent hierarchical structures directly in its unit structure (which is just a flat list). Instead, hierarchy is explicitly declared in feature-value pairs using the feature ‘sem-subunits’ on the semantic pole and the feature ‘syn-subunits’ on the syntactic pole. Since the ordering of units does not matter, FCG shares the advantage with dependency grammar formalisms that the same linguistic structures can be used for alternating word orders whereas strict constituency-based parsers have great difficulties in doing so.

On the relevance of FCG. Finally, I would like to comment on the ‘relevance’ of FCG for natural language processing (NLP). It is obvious that FCG was designed for artificial language evolution and not for NLP tasks so it does not have the same goals and objectives as existing parsers in computational linguistics. This becomes quickly apparent by looking at the scaffolds and abstractions of FCG: there is no need for word sense disambiguation, there is no scoping, etc. In short, FCG does not feature many of the solutions that computational linguists have found for solving complex problems in NLP tasks. In fact, FCG **only** contains mechanisms and processes which are necessary for the experiments on artificial language evolution.

In other words, the development of FCG is solely driven by the needs of the experiments and not by considerations of natural language problems. The relevance of FCG therefore cannot be found in parsing accuracy or performance, but rather on a more abstract level in the sense of design features for understanding open-ended communication and modeling the living and evolving aspects of grammar. Some of these design features can be of interest for natural languages as well, as I will explain in more detail in Chapter 8.

5.4 Parsing “Jack sweep dust off-floor”

With the meaning representation and FCG overview in mind, we can now dive into the details of linguistic processing in FCG. This section explains how the sentence *Jack sweep dust off- floor* is parsed, which is a simplification of the sentence *Jack sweeps the dust off the floor*. Again, this section does not provide a description of an actual English utterance, but a theoretical abstraction of one of the problems of argument realisation.

At the beginning of the parsing process, the hearer creates a first node in her reaction network, which is a coupled feature structure with an empty semantic and syntactic pole. I assume here that she is also capable of segmenting each morpheme or word into the separate strings “jack”, “sweep”, “dust”, “off-” and “floor”. These strings are all lumped together along with the observed word order (i.e. the ‘meets’-constraints) into the form-feature of one unit on the syntactic pole which I will call the top-unit. The label ‘top-unit’ is arbitrary but makes interpretation easier for human readers. On the semantic pole, the corresponding unit (also called ‘top-unit’) is still empty:

```
<Node-1:  coupled-feature-structure
((top-unit))
<==>
((top-unit
   (form ((string jack-unit ”jack”) (string sweep-unit ”sweep”)
          (string dust-unit ”dust”) (string off-unit ”off-”)
          (string floor-unit ”floor”) (meets jack-unit sweep-unit)
          (meets sweep-unit dust-unit) (meets dust-unit off-unit)
          (meets off-unit floor-unit)))))>
```

5.4.1 Unifying and merging lexical entries

In the next step, the hearer will perform a lexical look-up for all the words that she has put together in the syntactic/right pole of node-1. For this, we need a lexical entry for each of the words. The lexical entry for *jack* looks as follows:

```
<Lexical entry:  jack
((?top-unit
   (meaning (== (jack ?object-1))))
 ((J ?new-unit ?top-unit)
   (referent ?object-1)
   (sem-cat animate-object)))
<==>
((?top-unit
   (form (== (string ?new-unit ”jack”))))
 ((J ?new-unit ?top-unit)
   (syn-cat (== (pos noun)))))>
```

Note that the lexical entry for [jack] contains logic variables not only for the meaning but also for the unit-names. The unification engine of FCG can use these logic variables to match them against the unit-structure of the current node in the reaction network. Since the hearer acts in parsing mode, the right pole of the entry (under the directional marker <==>) needs to be unified with the right-pole of node-1. The right pole specifies that there has to be some unit (`?top-unit`) which must contain the feature `form`, which itself must contain in its value the feature-value `(string ?new-unit ”jack”)`. These constraints are indeed satisfied: the variable `?top-unit` can be bound to the unit `top-unit` in the current node because it fulfills all the necessary conditions.

Since unification is successful, the left pole, which contains the meaning of the lexical entry, can be merged with the left pole of node-1. The units that are marked with the J-operator were ignored during the unification phase, but are now integrated: the J-operator pulls the lexical information for *jack* down into a new unit and specifies that this new-unit is a sub-unit of the top-unit. The J-operator also merges additional features with this new unit concerning its referent and its semantic and syntactic categorisation. One could devise many other categorisations and features for *jack*, but they are not necessary for understanding the example here so they are left out for convenience's sake.

Since both unification and merge were successful, a new node is created in the reaction network. Here we see that the other words are still in the top-unit, but that there is a new unit for *jack* (which I conveniently label ‘jack-unit’ here but this may be any arbitrary symbol) both in the semantic and in the syntactic pole:

```
<Node-2:  coupled-feature-structure
((top-unit
   (sem-subunits (jack-unit)))
 (jack-unit
   (meaning ((jack ?object-1)))
   (referent ?object-1)
   (sem-cat animate-object)))
<==>
((top-unit
   (syn-subunits (jack-unit))
   (form ((string sweep-unit "sweep") (string dust-unit "dust")
          (string off-unit "off-") (string floor-unit "floor")
          (meets jack-unit sweep-unit) (meets sweep-unit dust-unit)
          (meets dust-unit off-unit) (meets off-unit floor-unit))))
 (jack-unit
   (form ((string jack-unit "jack")))
   (syn-cat ((pos noun)))))>
```

The lexical entries for *dust* and *floor* look almost exactly the same, apart from their semantic categorisation and meaning. Both entries also unify and merge successfully with the current node and license new nodes in the reaction network:

```
<Lexical entry:  dust
((?top-unit
   (meaning (== (dust ?object-2))))
 ((J ?new-unit ?top-unit)
   (referent ?object-2)
   (sem-cat moveable-object)))
<==>
((?top-unit
   (form (== (string ?new-unit "dust"))))
 ((J ?new-unit ?top-unit)
   (syn-cat (== (pos noun)))))>
```

```
<Lexical entry:  floor
((?top-unit
    (meaning (== (dust ?object-3))))
 ((J ?new-unit ?top-unit)
    (referent ?object-3)
    (sem-cat surface-object)))
<==>
((?top-unit
    (form (== (string ?new-unit ”floor”))))
 ((J ?new-unit ?top-unit)
    (syn-cat (== (pos noun)))))>
```

The lexical entry for *sweep*, however, is a bit more complicated. Its main function is the same as for the other lexical entries: given the string “sweep”, it will merge the meaning of this word with the semantic pole of the current node in the reaction network. The main difference with the other words lies in the semantic and syntactic categorisation of *sweep*. Take a look at the features ‘syn-frame’ in the syntactic pole and ‘sem-frame’ in the semantic pole:

```
<Lexical entry:  sweep
((?top-unit
    (meaning (== (sweep ?event-x)
                 (sweep-1 ?event-x ?obj-x)
                 (sweep-2 ?event-x ?obj-y)
                 (sweep-2 ?event-x ?obj-z))))
 ((J ?new-unit ?top-unit)
    (sem-frame (== (sem-role-agent ?unit-a ?obj-x)
                   (sem-role-surface ?unit-b ?obj-y)
                   (sem-role-moveable ?unit-c ?obj-y)
                   (sem-role-source ?unit-d ?obj-z)))))
<==>
((?top-unit
    (form (== (string ?new-unit ”sweep”))))
 ((J ?new-unit ?top-unit)
    (syn-cat (== (pos verb)))
    (syn-frame (== (syn-role-subject ?unit-1)
                   (syn-role-object ?unit-2)
                   (syn-role-oblique ?unit-3)))))>
```

At first glance, it seems that the lexical entry contains a predicate or case frame which lists the valence of a verb as in lexicalist approaches. The big difference is however that these frames do not directly list the argument structure of the verb but only its

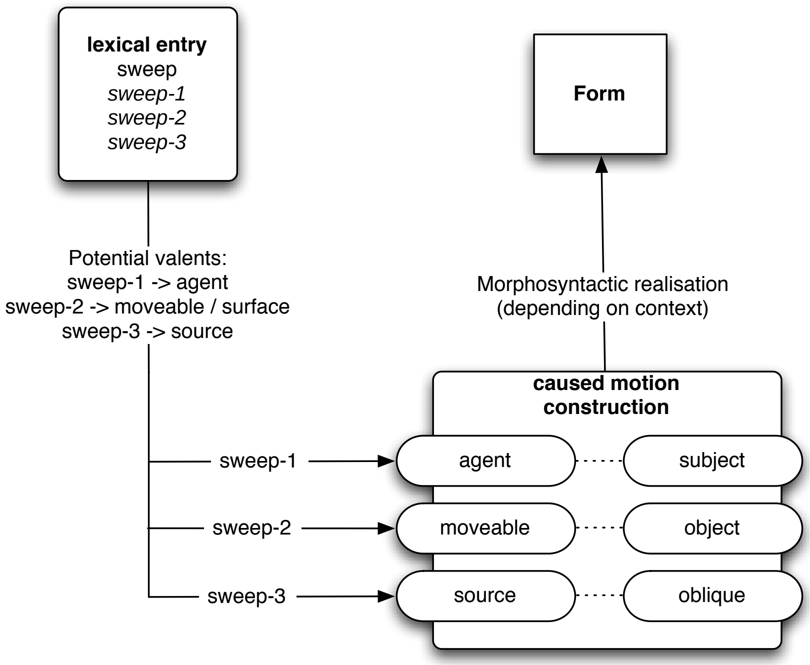


Figure 5.5: The relation between the meaning of an event and its morphosyntactic realisation is indirect and multilayered. This figure shows how a lexical entry introduces its 'potential valents' from which the construction selects the actual valency combination. The construction then maps this meaning onto a syntactic pattern, which in itself is realised as a certain form (depending on linguistic and extra-linguistic context)

potential valents. For example, the syn-frame in the syntactic frame only states that *sweep* can occur in an argument structure in which there may be some unit playing the role of subject, some unit playing the role of object, some unit playing the role of oblique, or several units playing a combination of these roles. None of these roles are however obligatory: the verb remains agnostic as to which of these valents actually should be realised or what the possible combinations may be. I will show later that it is the construction that will select from the potential valents what the **actual valency** of the verb is or will be in the utterance (also see Figure 5.5).

This architecture of potential valents is mirrored in the semantic frame. Here too, the verb only lists all its potential valents: there may be an agent role, a surface role, a moveable role, or a source role. The sem-frame does not specify which of these roles actually have to be realised, nor what the possible combinations are. It does specify, however, how these potential valences should be linked to the verb-specific participant roles in the meaning feature. For example, 'sem-role-agent' is linked to 'sweep-1' because they share the variable '?obj-x'. 'Sem-role-source' is linked to 'sweep-3' because they share the variable '?obj-z'. The participant role 'sweep-2' is even linked to two different potential valents: 'sem-role-surface' and 'sem-role-moveable'. The lexical entry thus allows participant roles to be mapped onto different semantic roles as I illustrated for the verb *sleep* in section 5.1 and for which I argued in section 4.2.4 of the previous chapter. Just as with the syntactic roles this mapping is selected by the construction in which the verb occurs.

Both the sem- and syn-frame also contain variables that have to be bound to the unit-names of the arguments to which the roles may be assigned later on. For example, the variable '?unit-1' should be bound to the unit that plays the role of subject (if present). Notice, however, that this variable is not the same one as the variable name of the unit that may play the agent role (if present): '?unit-a'. In most other grammar formalisms, such as HPSG, a direct link between 'agent' and 'subject' is assumed and alternating argument structures such as passives are *derived* through lexical rules. **I do not assume such a link between subject and agent**: having two different variable names reflects the fact that both can be bound to different units as is the case in the passive voice. I therefore consider the passive as an alternative argument structure construction instead of a derivational one. An active construction links agent and subject to each other by making their variables equal, whereas the passive construction features a different pattern by making other variables equal (e.g. those of the subject and the surface role as in *The floor was swept*). Here again, it is the construction that decides and not the verb. I will return to this matter in section 8.2.3 of Chapter 8.

The following chapters in Part II will demonstrate how these potential valents can be gradually acquired and constructed through language use. They should therefore not be seen as a rigid set of possibilities or as some set of innate categories. Instead the potential uses of a verb can be extended (and shrunk) if needed for communicative purposes, and each possibility may become conventionalised or become obsolete in the speech community.

Unifying and merging the lexical entries for *sweep*, *floor* and *dust* (the ordering does not matter) will lead the hearer to a fifth node in the reaction network:

```
<Node-5:  coupled-feature-structure
((top-unit
   (sem-subunits (jack-unit sweep-unit dust-unit floor-unit)))
 (jack-unit
   (meaning ((jack ?object-1)))
   (referent ?object-1)
   (sem-cat animate-object))
 (sweep-unit
   (meaning ((sweep ?event-x)
             (sweep-1 ?event-x ?obj-x)
```

```
              (sweep-2 ?event-x ?obj-y)
              (sweep-3 ?event-x ?obj-z)))
    (sem-frame ((sem-role-agent ?unit-a ?obj-x)
                (sem-role-surface ?unit-b ?obj-y)
                (sem-role-moveable ?unit-c ?obj-y)
                (sem-role-source ?unit-d ?obj-z))))
  (dust-unit
    (meaning ((dust ?object-2)))
    (referent ?object-2)
    (sem-cat moveable-object))
  (floor-unit
    (meaning ((floor ?object-3)))
    (referent ?object-3)
    (sem-cat surface-object)))
<==>
((top-unit
    (syn-subunits (jack-unit sweep-unit dust-unit floor-unit))
    (form ((string off-unit "off-") (meets jack-unit sweep-unit)
           (meets sweep-unit dust-unit) (meets dust-unit off-unit)
           (meets off-unit floor-unit))))
  (jack-unit
    (form ((string jack-unit "jack")))
    (syn-cat ((pos noun))))
  (sweep-unit
    (form ((string sweep-unit "sweep")))
    (syn-cat ((pos verb)))
    (syn-frame ((syn-role-subject ?unit-1)
                (syn-role-object ?unit-2)
                (syn-role-oblique ?unit-3))))
  (dust-unit
    (form ((string dust-unit "dust")))
    (syn-cat ((pos noun))))
  (floor-unit
    (form ((string floor-unit "floor")))
    (syn-cat ((pos noun)))))>
```

The meanings in the above coupled feature structure are however still unlinked (see section 5.2). In other words, the hearer knows at this stage the meaning of the individual words, but not who is doing what in the sweep-event: the variables that accompany the meaning of *sweep* (‘?obj-x’, ‘?obj-y’ and ‘?obj-z’) are not shared by any of the arguments (‘?object-1’, ‘?object-2’ and ‘?object-3’). We therefore need to unify and merge the correct construction that is able to detect the unresolved variable equalities and do something about it.

5.4.2 A syntactic case marker

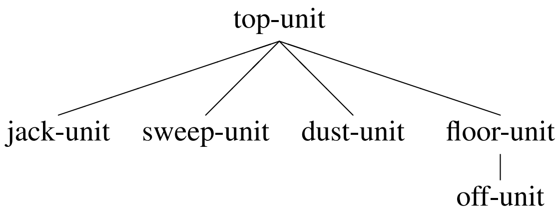
There is still the string “off-” left in the top-unit that needs to be parsed before the construction can be applied. In this example I analyse *off-* as some kind of simple case marker which assigns the oblique case to the argument that immediately follows it. By treating *off-* as a case marker I can immediately illustrate how markers are represented in the experiments as well.

In line with its definition in section 4.2.5, a syntactic case marker is dissociated from a particular semantic role. I therefore implement case markers in a morphological rule or construction (a ‘morph-rule’) in which both the left and the right pole are syntactic (so both poles operate on the syntactic pole of the current node in the reaction network). One could say in this approach that a case marker has a syntactic or grammatical meaning rather than a semantic one. The notion of potential valents can also be applied to capture the polysemous nature of case markers, but for simplicity’s sake I will assume here that there is a one-to-one mapping between the syntactic role ‘oblique’ and the marker *off-*.

The morph-rule specifies that, during parsing, there needs to be a unit which contains in its form-feature the string “off-” and a word order constraint that says that the marker immediately precedes (‘meets’) some other unit. If this unifies (and it does with the right pole of node-5), the left pole merges with the syntactic pole of the current node in the reaction network. The information added here is that the role of oblique is assigned to the other unit (which immediately followed the marker). Note that this other unit should already be present in the current node as a sub-unit of the top-unit. This is indeed the case: the variable ‘?some-unit’ can be bound to ‘floor-unit’ which was already present after unifying and merging the lexical entry of *floor*.

The morph-rule also contains a two-legged operation using the J-operator: in a first step, a new unit is created for *off-*, and in a second step the J-operator specifies that the newly-made unit must become a sub-unit of the other unit that immediately followed it (floor-unit). Making the marker immediately dependent of floor-unit has some advantages that are known from dependency grammars such as the fact that there is no need for a specific head-daughter feature or some additional abstract constituent structure (Nivre, 2006). The two-legged operation proposed here really stretches the use of the J-operator and further investigation is needed to see whether FCG could feature more general-purpose structure building or dependency operators.

Without repeating the entire coupled feature structure, the syntactic structure now looks as follows:



The morphological rule itself looks as follows:

```
<Morph-rule:  off-
((?top-unit
   (syn-subunits (== ?some-unit)))
 (?some-unit
   (syn-role syn-role-oblique)))
<==>
((?top-unit
   (form (== (string ?marker-unit ”off-”)
             (meets ?marker-unit ?some-unit))))
 ((J ?marker-unit ?top-unit))
 ((J ?some-unit ?top-unit (?marker-unit))))>
```

5.4.3 The caused motion construction

The utterance *Jack sweeps the dust off the floor* is a typical example of the caused motion construction (Goldberg, 1995, chapter 7), which here carries the meaning of ‘X causes Y to move Z by sweeping’. In the semantic pole, the construction selects from the verb the semantic roles of agent, moveable patient and (in this simplified grammar) source. On the syntactic pole it assigns the syntactic roles of subject, object and oblique to the arguments. The construction looks as follows:

```
<Construction:  caused-motion
((?top-unit
   (sem-subunits (== ?unit-a ?unit-b ?unit-c ?unit-d)))
 (?unit-a
   (sem-frame (== (sem-role-agent ?unit-b ?obj-x)
                  (sem-role-moveable ?unit-c ?obj-y)
                  (sem-role-source ?unit-d ?obj-z))))
 (?unit-b
   (referent ?obj-x)
   (sem-cat animate-object))
 (?unit-c
   (referent ?obj-y)
   (sem-cat moveable-object))
```

```
 (?unit-d
   (referent ?obj-z)))
<==>
((?top-unit
   (syn-subunits (== ?unit-a ?unit-b ?unit-c ?unit-d))
   (form ((meets ?unit-b ?unit-a) (meets ?unit-a ?unit-c)
          (meets ?unit-c ?unit-d))))
 (?unit-a
   (syn-cat (== (pos verb)))
   (syn-frame (== (syn-role-subject ?unit-b)
                  (syn-role-object ?unit-c)
                  (syn-role-oblique ?unit-d))))
 (?unit-d
   (syn-role syn-role-oblique))
 ((J ?unit-b NIL)
   (syn-role syn-role-subject))
 ((J ?unit-c NIL)
   (syn-role syn-role-object)))>
```

Since the hearer is parsing, the right pole needs to be unified with the current node in the reaction network. The right pole here demands there to be some unit with at least four sub-units and with a certain word order among them (i.e. the ‘meets’-constraints). The last argument also has to have the syntactic role of oblique. The right pole of the current node satisfies all the constraints: jack-unit receives subject status and dust-unit becomes the object (which is here taken care of by the J-operator). The unit for *floor* was already marked as oblique by the morphological rule that was shown in the previous section.

Thanks to the word-order constraints, the construction can unambiguously bind the variables ‘?unit-b’ to ‘jack-unit’, ‘?unit-c’ to ‘dust-unit’ and ‘?unit-d’ to ‘floor-unit’. The construction then links the syntactic roles to the semantic roles by making the necessary variables equal: the subject jack-unit is assigned the role of agent, the object dust-unit is assigned the role of moveable (object) and the oblique floor-unit is assigned the role of source. By doing so, the construction can also link the meanings to each other: sem-role-agent shares the variable ‘?obj-x’ with sweep-1 and the referent of jack-unit, sem-role-moveable shares the variable ‘?obj-y’ with sweep-2 and the referent of dust-unit, and sem-role-source shares the variable ‘?obj-z’ with sweep-3 and the referent of floor-unit. This leads to the following node in the reaction network:

```
<Node-7:  coupled-feature-structure
((top-unit
   (sem-subunits (jack-unit sweep-unit dust-unit floor-unit)))
 (jack-unit
   (meaning ((jack ?obj-x)))
```

```
   (referent ?obj-x)
   (sem-role sem-role-agent)
   (sem-cat animate-object))
 (sweep-unit
   (meaning ((sweep ?event-x)
             (sweep-1 ?event-x ?obj-x)
             (sweep-2 ?event-x ?obj-y)
             (sweep-3 ?event-x ?obj-z)))
   (sem-frame ((sem-role-agent jack-unit ?obj-x)
               (sem-role-surface dust-unit ?obj-y)
               (sem-role-moveable dust-unit ?obj-y)
               (sem-role-source floor-unit ?obj-z))))
 (dust-unit
   (meaning ((dust ?obj-y)))
   (referent ?obj-y)
   (sem-role sem-role-moveable)
   (sem-cat moveable-object))
 (floor-unit
   (meaning ((floor ?obj-z)))
   (referent ?obj-z)
   (sem-role sem-role-source)
   (sem-cat surface-object)))
<==>
((top-unit
   (syn-subunits (jack-unit sweep-unit dust-unit floor-unit))
   (form ((meets jack-unit sweep-unit) (meets sweep-unit dust-unit)
          (meets dust-unit off-unit))))
 (jack-unit
   (form ((string jack-unit ”jack”)))
   (syn-role syn-role-subject)
   (syn-cat ((pos noun))))
 (sweep-unit
   (form ((string sweep-unit ”sweep”)))
   (syn-cat ((pos verb)))
   (syn-frame ((syn-role-subject jack-unit)
               (syn-role-object dust-unit)
               (syn-role-oblique foor-unit))))
 (dust-unit
   (form ((string dust-unit ”dust”)))
   (syn-role syn-role-object)
   (syn-cat ((pos noun))))
 (floor-unit
   (syn-subunits (off-unit))
   (form ((string floor-unit ”floor”)))
```

```
   (syn-role syn-role-oblique)
   (syn-cat ((pos noun))))
 (off-unit
   (form ((string off-unit "off-") (meets off-unit floor-unit)))))>
```

The hearer has now used all the matching linguistic items that she knows, so now she can extract the meanings from the semantic pole of the coupled feature structure. This yields the following meaning:

(5.19) $\exists$?obj-x, ?obj-y, ?obj-z, ?event-x: jack(?obj-x) $\wedge$ dust(?obj-y) $\wedge$ floor(?obj-z) $\wedge$ sweep(?event-x) $\wedge$ sweep-1(?event-x , ?obj-x) $\wedge$ sweep-2(?event-x, ?obj-y) $\wedge$ sweep-3(?event-x, ?obj-z)

As can be seen in the meaning, all variables that refer to the same referent have been made equal by the construction. The hearer thus knows that *jack* was the sweeper, that *dust* was the thing being swept, and that the *floor* was the source of the motion. In the experiments reported in this thesis, the agents will then match this parsed meaning against their world model.

5.5 Producing “jack sweep floor”

This section gives an overview of how an utterance such as *jack sweep floor* can be produced in Fluid Construction Grammar. In this case, the caused motion construction is not used, but a construction which maps the semantic frame ‘X acts on surface Y’ onto the syntactic frame ‘Subject-Verb-Object’. The speaker starts with the following meaning (in which NIL stands for ‘empty’ or ‘not profiled’):

(5.20) ((jack object-1) (floor object-2) (sweep event-1) (sweep-1 event-1 object-1) (sweep-2 event-1 object-2) (sweep-3 event-1 NIL))

In order to verbalise this meaning, the speaker constructs a reaction network. The first node in the network is a coupled feature structure in which the entire meaning is placed into one unit in the semantic pole, and in which the syntactic pole is still empty:

```
<Node-1:  coupled-feature-structure
((top-unit
   (meaning ((jack object-1) (floor object-2)
            (sweep event-1) (sweep-1 event-1 object-1)
            (sweep-2 event-1 object-2) (sweep-3 event-1 NIL)))))
<==>
((top-unit))>
```

5.5.1 Unifying and merging lexical entries

Next, the speaker builds new nodes in the network by unifying and merging the lexical entries that cover these meanings. Suppose that the speaker has the same lexical entries as given in the previous section and that all three of them (*jack, sweep* and *floor)* are a successful match, then the new coupled feature structure looks as follows:

```
<Node-4:  coupled-feature-structure
((top-unit
   (sem-subunits (jack-unit sweep-unit floor-unit)))
 (jack-unit
   (meaning ((jack object-1)))
   (referent object-1)
   (sem-cat animate-object))
 (sweep-unit
   (meaning ((sweep event-1)
             (sweep-1 event-1 object-1)
             (sweep-2 event-1 object-2)
             (sweep-3 event-1 NIL)))
   (sem-frame ((sem-role-agent ?unit-a object-1)
               (sem-role-surface ?unit-b object-2)
               (sem-role-moveable ?unit-c object-2)
               (sem-role-source ?unit-d NIL))))
 (floor-unit
   (meaning ((floor object-2)))
   (referent object-2)
   (sem-cat surface-object)))
<==>
((top-unit
   (syn-subunits (jack-unit sweep-unit floor-unit)))
 (jack-unit
   (form ((string jack-unit "jack")))
   (syn-cat ((pos noun))))
 (sweep-unit
   (form ((string sweep-unit "sweep")))
   (syn-cat ((pos verb)))
   (syn-frame ((syn-role-subject ?unit-1)
               (syn-role-object ?unit-2)
               (syn-role-oblique ?unit-3))))
 (floor-unit
   (form ((string floor-unit "floor")))
   (syn-cat ((pos noun)))))>
```

Since the speaker knows which meaning she wants to express, there are no unresolved variable equalities in the meanings in the semantic pole. However, so far no semantic or syntactic roles have been assigned yet: the lexical entry of *sweep* has merely introduced its potential valents, but not its actual valency. This is reflected by the fact that the arguments do not have the feature 'sem-role' or 'syn-role' yet and that there are still variables left for the unit-names in both the syn- and sem-frame. It is also still undecided which participant should be mapped onto subject and which onto object or oblique.

5.5.2 The agent-acts-on-surface construction

Trying to unify the caused motion construction leads to a failure: first of all, *floor* is categorised as a static surface-object and thus violates the construction's constraint that the patient has to be moveable, and second, the source argument is missing. So a different construction needs to be unified and merged to travel to the next node in the network. The following 'agent-acts-on-surface' construction would do the trick:

```
<Construction:  agent-acts-on-surface
((?top-unit
   (sem-subunits (== ?unit-a ?unit-b ?unit-c)))
 (?unit-a
   (sem-frame (== (sem-role-agent ?unit-b ?obj-x)
                  (sem-role-surface ?unit-c ?obj-y))))
 (?unit-b
   (referent ?obj-x)
   (sem-cat animate-object))
 (?unit-c
   (referent ?obj-y)
   (sem-cat surface-object)))
<==>
((?top-unit
   (syn-subunits (== ?unit-a ?unit-b ?unit-c ?unit-d))
   (form ((meets ?unit-b ?unit-a) (meets ?unit-a ?unit-c))))
 (?unit-a
   (syn-cat (== (pos verb)))
   (syn-frame (== (syn-role-subject ?unit-b)
                  (syn-role-object ?unit-c))))
 ((J ?unit-b NIL)
   (syn-role syn-role-subject))
 ((J ?unit-c NIL)
   (syn-role syn-role-object)))>
```

First the speaker tries to unify the semantic pole with that of the current node in the reaction network. This is a success because all the constraints are satisfied: the construction expects some unit with three sub-units, one of which is an animate-object and one of which is a surface-object. The construction also selects the semantic roles of ‘agent’ and ‘surface’ from the verb’s potential valents and assigns them to the correct arguments. Next, the right pole of the construction is merged with the right pole of the current node. Since it is specified that the agent maps onto subject and the surface maps onto object in this construction, the correct syntactic roles are assigned to the arguments, including their word order. The new node looks as follows:

```
<Node-5:  coupled-feature-structure
((top-unit
   (sem-subunits (jack-unit sweep-unit floor-unit)))
 (jack-unit
   (meaning ((jack object-1)))
   (referent object-1)
   (sem-role sem-role-agent)
   (sem-cat animate-object))
 (sweep-unit
   (meaning ((sweep event-1)
             (sweep-1 event-1 object-1)
             (sweep-2 event-1 object-2)
             (sweep-3 event-1 NIL)))
   (sem-frame ((sem-role-agent jack-unit object-1)
               (sem-role-surface floor-unit object-2)
               (sem-role-moveable floor-unit object-2)
               (sem-role-source ?unit-d NIL))))
 (floor-unit
   (meaning ((floor object-2)))
   (referent object-2)
   (sem-role sem-role-surface)
   (sem-cat surface-object)))
<==>
((top-unit
   (syn-subunits (jack-unit sweep-unit floor-unit))
   (form ((meets jack-unit sweep-unit)
          (meets sweep-unit floor-unit))))
 (jack-unit
   (form ((string jack-unit ”jack”)))
   (syn-role syn-role-subject)
   (syn-cat ((pos noun))))
 (sweep-unit
   (form ((string sweep-unit ”sweep”)))
   (syn-cat ((pos verb)))
```

```
   (syn-frame ((syn-role-subject jack-unit)
              (syn-role-object floor-unit)
              (syn-role-oblique ?unit-3))))
 (floor-unit
   (form ((string floor-unit "floor")))
   (syn-role syn-role-object)
   (syn-cat ((pos noun)))))>
```

The speaker has no other constructions or linguistic items that could unify and merge, so she is almost done processing. In the final phase, she takes all the formal constraints specified in the syntactic pole of the last node and renders them into the utterance *jack sweep floor*.

5.6 Networks and conventionalisation

The previous two sections showed how lexical and constructional meanings can integrate through the potential valents of the verb on the one hand, and the selection of the actual valency by the constructions on the other. I also suggested that the list of potential valents can be extended if needed for communication. However, this solution is too powerful because it does not explain why speakers of English prefer not to use *sweep* in for example the ditransitive construction as in **he swept him the dust*. We therefore need an additional strategy to restrict the powers of the proposed representation.

As said before, utterances such as **he swept him the dust* are perfectly intelligible but somehow speakers (usually) refrain from using words in constructions that they are not conventionally associated with. Convention or entrenchment can intuitively be captured in a network as illustrated in Figure 5.6. The basic idea is the following: we never observe words in total isolation but always in language use. We can therefore keep links between constructions that reflect their past co-occurrences. For example, *sweep* has a link to the intransitive construction, the caused motion construction, etc. The link reflects conventionalisation through a confidence score: the higher this score, the more confident the speaker feels integrating the lexical entry with the construction. The lower the score, the more 'strain' there is to go ahead and fuse the lexical entry with the construction. Scores in the network can be changed each time the lexical entry and the construction co-occur: the score increases if they are used in successful communication, but decreases if co-occurrence leads to communicative failure. Each interaction also yields the possibility of adding new links. I will come back to the exact nature of this network in the following chapters.

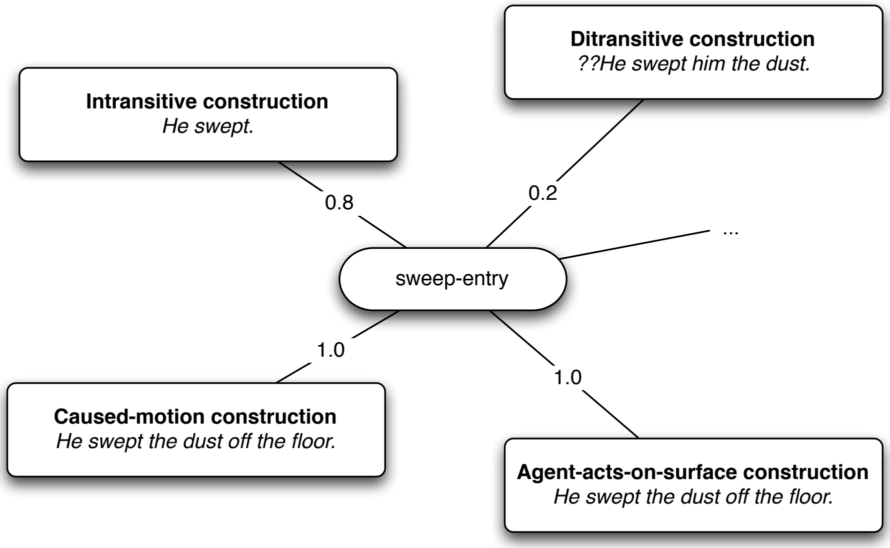


Figure 5.6: Lexical entries keep a link to all the constructions that they integrated with. Each link has a specific confidence score which reflects how confident the language user is that the lexical entry can be fused with the construction. The higher the score, the more conventionalised the link is. The lower the score, the more 'strain' the speaker will have to use the entry and the construction together. This network thus captures conventionalised patterns in a language.

5.7 Summary: potential valents and actual valency

This chapter presented a fully operational implementation in Fluid Construction Grammar of how the semantic representation of a verb can interact with argument structure constructions in order to yield different patterns of argument realisation. After reviewing some problems with the definition of this 'fusion process' offered by Goldberg (1995, p. 49), I proposed that the lexical entry of a verb introduces its potential semantic and syntactic roles (its 'potential valents') to the structure which is being produced or parsed by a language user. This 'potential' – which is gradually acquired and possibly even extended through language use – does not provide a priori combinations, nor does it specify how the meanings of an utterance should be combined with each other. In the next step it is the argument structure construction that selects how this potential should be used for mapping meaning to form and vice versa. In other words, the construction decides on the actual valency and argument structure.

This proposal was illustrated with the production and parsing of two example utterances: *jack sweep dust off-floor* and *jack sweep floor*. Both examples are simplified representations of English-like sentences and are directly related to experiments reported in the rest of Part II. As such, they can be regarded as 'design solutions' for the problem of argument realisation which have to be complemented in linguistic theories with constructions for tense, aspect, etc. The traces of the examples show how the licensing of a construct depends on a combination of all constructions and lexical items involved rather than specifying all constraints in either the lexical entry of the verb or in various linking rules (as is common practice in lexical approaches). Finally, I suggested that co-occurrences between constructions should be captured in a network that reflects the degrees of conventionalisation of argument realisation patterns.

Chapter 6

Baseline Experiments

The grammatical square (see Figure 4.6) illustrates the multifunctional and indirect nature of case marking. It can also be read as a 'research roadmap' for experiments on the formation of a grammar for case. More specifically, the experiments must investigate how a population of agents can evolve a grammar which takes care of (1) the mapping between event-specific participant roles and generalised semantic roles, (2) the mapping between semantic roles and grammatical cases, and (3) the mapping between cases and surface case markers. Constructing and aligning this kind of grammar in a multi-agent population is incredibly difficult because all mappings are agent-internal and are thus not directly observable by other agents. Moreover, these mappings vary depending on the communicative and linguistic context so the agents have to be capable of dealing with polysemy. Finally, linguistic conventions may constantly change so the agents have to be able to adapt to newly propagated constructions.

In this chapter I will present three baseline experiments which first replicate simulations reported by Steels (2002b, 2004c) and then lift them towards a multi-agent simulation. I will specifically focus on which diagnostics, repairs and alignment strategies are needed for each step of increased complexity. In the next section, I will first give an overview of the experimental set-up which is shared by all experiments after which the simulations themselves are reported.

6.1 Experimental set-up

The simulations in this thesis investigate how a population of artificial agents can autonomously construct a shared grammatical system for marking event structure (in the form of a case grammar). As argued in section 3.2, we need to hypothesise which cognitive mechanisms and external pressures are the minimal ingredients that enable the agents to do so, and we need to operationalise them into computational processes and a world environment. Since case marking is a complex grammatical phenomenon, I will divide the topic into several subparts which follow the stages of development that were identified in section 4.2.

	Detecting and resolving variable equalities	Invention and adoption of new markers	Reuse and generalisation of existing markers
Baseline experiment 1	+	-	-
Baseline experiment 2	+	+	-
Baseline experiment 3	+	+	+

Table 6.1: This table shows the key cognitive abilities which are given to the agents in the baseline experiments. In the first experiment, agents are 'intelligent' enough to resolve variable equalities by using their world model but they cannot express these equalities in their language. In experiment 2, the agents can decide to invent new markers to indicate the linking of equal variables but they cannot generalise or abstract. In the third experiment, the agents can perform analogical reasoning over event structures to generalise existing markers.

6.1.1 Key cognitive abilities and self-assessment criteria

In the problem-solving approach adopted in this thesis (see section 3.1.3), innovation and language change happens in three steps: (1) the speaker innovates and is therefore the main cause of potential language change, (2) the hearer tries to infer and learn the innovation through an 'abductive process' (as opposed to uniquely relying on induction or genetically evolved innate knowledge), and (3) the innovation propagates in the population. Step (1) may be skipped when the hearer overgeneralises or imposes more systematicity on the utterance than introduced by the speaker. Hearer-based innovation or 'reanalysis' can be described in terms of the same cognitive processes as involved in speaker-based innovation so the two sources of innovation are complementary to each other (Hoefler & Smith, 2008). Step (3) implies that an innovation only becomes a linguistic convention if it has been adopted by a sufficient number of language users. Propagation is made possible through the alignment strategies of the agents.

This approach requires a population of agents endowed with rich cognitive capabilities. As argued in section 3.2, agents can only make use of local measures such as communicative success and cognitive effort for steering their linguistic behaviour. The cognitive mechanisms of the agents therefore have to enable the agents to move into a meta-level in which they can **self-assess** what problems they encounter during communication, whether there is opportunity for optimisation, or what the reasons for success and failure are in a language game. They need to be able to couple this **diagnosis** to **repair strategies** in order to solve their communicative problems and to their **alignment strategies** in order to converge on a shared language. All the experiments in this thesis mainly focus on the effect of these three aspects: diagnostics, repairs, and alignment strategies.

The baseline experiments reported in this chapter first of all replicate the two-agent simulations reported by Steels (2002b, 2004c). These experiments only focus on the first three stages of case marker development ranging from no marking to the formation and marking of semantic roles. Additionally, the experiments are pushed forward to multi-agent simulations in which language convergence becomes the main issue. All three baseline experiments share the same world environment, communicative task, and assumptions and scaffolds (see below). However, they differ in the **key cognitive abilities** which the agents are endowed with. I define ‘key cognitive abilities’ as those mechanisms which are hypothesised to be crucial for the transition in the grammar from one stage to the next, and of which the simulations need to demonstrate or falsify whether this is indeed the case. Table 6.1 illustrates the difference between the baseline experiments in terms of key cognitive abilities and can be summarised as follows:

1. In baseline experiment 1 (corresponding to stage 1 – section 4.2.2), agents are given the diagnostic to figure out how the meanings of lexical items are linked to each other by exploiting the situatedness of the interaction. However, the agents have no means of extending their language to explicitly mark the relations between words;

2. In baseline experiment 2 (corresponding to stage 2 – section 4.2.3), agents are endowed with a repair strategy which enables them to invent a participant role-specific case marker for optimising communication;

3. In baseline experiment 3 (corresponding to stage 3 – section 4.2.4), agents are endowed with the capacity to perform analogical reasoning over event structures. They can exploit this capacity for generalising existing markers to cover new participant roles. As I will show later, generalisation is not a goal in itself but rather a side-effect of the need for optimising communication.

These key cognitive abilities will be explained in more detailed along with the experiments further down in this chapter. The abilities are each time given and fixed by the experimenter and the simulations do not explain where they come from. However, the global vision underlying this work is that speakers and hearers can autonomously configure their language capacity by recruiting cognitive mechanisms that are also used for other tasks such as hierarchy building operators (Steels, 2007, also see section

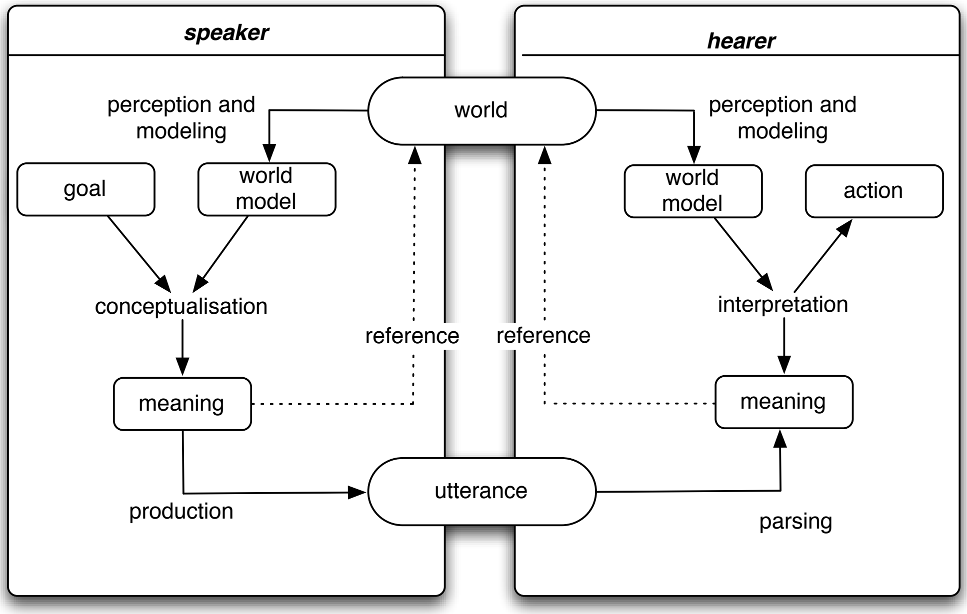


Figure 6.1: The semiotic cycle involved in language games. There are three systems working in tight interaction with each other: the sensory-motor system (perception and modeling), the conceptual/intentional system (conceptualisation and interpretation), and the linguistic system (production and parsing). This thesis mainly focuses on the latter one and is complementary to other research efforts on grounded language use and conceptualisation.

2.5.3). This process is driven by needs in communicative success, expressive power and the conventions adopted by other agents in the population. Here again, agents have to be capable of self-assessing when to recruit a mechanism, which ones are the best candidates, and to self-regulate the semantic complexity of their communication. For simulations that investigate self-regulation and reconfiguration of the language faculty on the longer run, see Steels & Wellens (2006, 2007). In the remainder of this section, I will describe those aspects of the simulations which are shared by all the experiments.

6.1.2 Description games

An obvious requirement for developing a grammar for marking event structure is **communication about events**. This is operationalised in the form of the **description game**, a routinised communicative interaction which involves the complete semiotic cycle as illustrated in Figure 6.1. Applied to the simulations in this thesis, the interaction pattern conforms to the following script:

1. Two agents are randomly selected from the population. One will act as speaker, the other as hearer. The speaker and hearer start a **local language game** so the other agents cannot observe it;
2. **Joint attention** (Tomasello, 1995) between the agents is required and assumed. This is operationalised by giving the agents a shared context. The context contains one or more events (depending on the complexity of the game) which are observed by the agents;
3. Both the speaker and the hearer build a **world model** based on the observed events. The world model consists of a series of facts in the memory;
4. The speaker is given a communicative goal. In this thesis, the goal is always to **make an assertion** about a certain state of affairs in the context (i.e. the speaker has to describe an event);
5. The speaker chooses an event to describe and **conceptualises** a meaning for it. Conceptualisation involves profiling of the event and finding a discriminating meaning for the participants in the event (see further below);
6. The speaker then verbalises the meaning by **producing** an utterance which is transmitted to the hearer;
7. The hearer observes the utterance and **parses** it;
8. The hearer then **interprets** the parsed meaning by comparing it to the facts in the world model. This leads to the **mental action** of agreeing or disagreeing with the description;
9. The hearer signals agreement with the description if the parsed meaning is unambiguously compatible with the hearer's world model, or signals disagreement if it is not. Agreement means **communicative success**, disagreement means **communicative failure**. No other kinds of feedback or requests for information are included;
10. Based on the outcome of the game, both agents **consolidate** their linguistic inventories.

6.1.3 The world, sensory-motor input and conceptualisation

The semiotic cycle in Figure 6.1 clearly shows that linguistic processing is only one part of the general cognitive architecture which is involved in communication. Even though the experiments in this thesis mainly focus on the production and parsing of linguistic utterances, at least two other systems are directly relevant for communication: the sensory-motor system (in a broad sense responsible for dealing with the sensory experience and for building a world model) and the conceptual-intentional system (responsible for conceptualisation and interpretation as well as concept formation). These

Figure 6.2: The world environment consists of dynamic scenes in which puppets perform various actions such as walking and pushing blocks to each other. Here, one puppet 'gives' a block to another puppet by sliding it over a table. The scene contains four participants (two puppets, a block and a table), a 'ground' and various micro- and macro-events.

three 'systems' work together in tight interaction and without a clear-cut division between them.

In this section, I will describe what kind of world environment is used in the experiments, what kind of world model is built by the sensory-motor system, and what kind of meanings are conceptualised for communication. These three elements remain constant and are shared by all the experiments.

The environment. The environment in the experiments consists of dynamic real-world scenes from a small puppet theater. The puppets perform various actions such as moving, disappearing from a scene, walking towards each other, and carrying objects. The use of real-world scenes is part of earlier work on grounded language communication (Steels, 2002b; Steels & Baillie, 2003) and is not essential to the dynamics of the models reported in this thesis: a simulated world suffices for the scope of this dissertation and can in fact be used for scaling up the experiments to larger worlds. The choice for using data from real-world scenes was made in order to demonstrate that the models *can* be incorporated into research on the grounding of communication.

For the baseline experiments, 20 different scenes were recorded comprising 207 event tokens (belonging to 15 different event types, see Baillie, 2002). There are 103 event tokens which take one participant (e.g. a 'move-event'), 99 event tokens which potentially take two participants (e.g. a 'walk-to'-event) and 5 event tokens which potentially involve three participants (e.g. a 'give-event'). Given the conceptualisation algorithm (see below), this leads to a frequency of about 64% of utterances involving one participant, about 34% of utterances involving two participants and only 2% of utterances involving three participants. In most experiments each event type was given the same frequency in order to balance this skewed frequency. As I will demonstrate later, equal frequency offers a much clearer picture of the propagation and convergence dynamics in the experiments.

Sensory-motor input. The real-world scenes were recorded using two SONY pan-tilt cameras (EVI-D31) which were hooked up to computers running the PERACT system (Baillie, 2002; Baillie & Ganascia, 2000; Steels & Baillie, 2003), which was designed for the visual recognition of events and which is related to other attempts in visual-recognition such as Siskind (2000). Even though two cameras were used, this thesis uses only the data obtained through one of them. The simulations are thus not affected by differences in world models due to noisy recognition of the events or differences in visual perspective. These difficulties are very interesting for investigating the robustness of the model in grounded communication, but are not part of this thesis.

Since the vision system is quite complex (but fortunately well-documented in Baillie, 2002), I will restrict my discussion to its general architecture and to the design choices that are important for understanding what information is delivered to the language system. The PERACT system first delineates objects based on colour histograms and then groups the pixels belonging to the same object together. These objects (two puppets, a table, two blocks, etc.) were learned in advance and PERACT can handle seven of them simultaneously in one scene. Starting from basic visual 'primitives' or 'micro-events' such as touching, movement and appearance, the system tracks the objects in real-time and assembles more complex descriptions (or 'macro-events') when it recognises a pattern in the scene. This recognition is often unreliable, but saliency, confidence and hierarchy are used to categorise the scene in terms of micro- and macro-events. This categorisation of events is then supplied to the linguistic system.

For each agent, the filtered results of sensory processing is then represented as a series of facts in the memory. For example, an event in which one of the puppets 'moves inside a yellow house' is represented as follows:

(6.1)
```
(move-inside event-163190 true)
(move-inside-1 event-163190 jill)
(move-inside-2 event-163190 house-1)
(girl jill) (jill jill) (house house-1)
(yellow house-1) (boy jack) (jack jack)
```

As can be seen in example 6.5, there are three objects in the scene (two puppets called 'Jack' and 'Jill', and a house), but only two of them are participants in the move-inside-event ('Jill' and the house). For each object, the vision system delivers at least two facts which can be used during conceptualisation for discriminating the objects from the other ones in the same scene (see below). These facts are very simple (for example 'house' and 'yellow' for the house-object), but can be interpreted in a more general sense as being the features describing an object (e.g. in a more detailed implementation, the feature-values of a R(ed)G(reen)B(lue)-channel could be used instead of the feature 'yellow').

The labels of these facts thus carry English names, **but should not be interpreted as such**. For example, 'move-inside' actually involves a puppet moving behind another object. The vision system, which is based on colour recognition, does not have any notion of a container or some kind of 'insideness'. Instead it sees one colour blob (the girl puppet) moving towards another one (the yellow 'house') until the colour regions 'touch' each other after which the girl puppet disappears out of sight. So the label 'move-inside' does not reflect what actually happens in the scene (in terms of human conceptualisation) but it is used for facilitating the analysis of the event recognition by the experimenter.

From the above it should also be clear that the PERACT system describes dynamic scenes in different levels of complexity. For example, the move-inside-event is itself a macro-event which can be decomposed into sub-events (which are macro-events themselves or micro-events, and which are also represented as facts in the memories of the agents). The micro-events for a move-inside-event are primitive relations such as 'visible', 'distance-decreasing', 'movement', 'touching' and 'disappearing'. The PERACT system offers a hierarchical description of events including time stamps for their beginnings and ends. As I will explain in section 6.4, the simulations reported in this thesis disregard this temporal and hierarchical information and treat the event structure of an event as a flat list of micro-events. This choice was made in order to focus on the participant roles as a whole rather than on causal-aspectual parts of the event-structure. I will come back to this choice when discussing the experiments on Stage IV in the development of case markers in section 7.5.

One final remark regarding the sensory-motor input has to do with the status of the visual 'primitives'. I do not make any claims about whether they are innate or not, nor do I claim that they represent a realistic set for categorising events in human cognition (neither in terms of size nor in terms of quality). The idea is rather that events can be decomposed into a much richer representation that allows analogical reasoning and the comparison of event structures. As discussed in section 4.1.2, event structure and event decomposition now plays a key role in many linguistic theories of argument realisation.

Conceptualisation. The categorisation of the scenes in terms of event types and objects (and their features) is already taken care of by the sensory-motor system. Conceptualisation in the simulations is therefore a very basic, but nonetheless crucial operation. First of all, during conceptualisation the speaker agent decides on the **event profile** that it wants to express. 'Event profile' should be interpreted in roughly the same way as is done by Croft (1998, see section 4.1.2) and most approaches in cognitive linguistics. Since the agents do not take temporal-aspectual information of the events into account, profiling events essentially consists of deciding which participants have to be expressed explicitly. For the aforementioned move-inside-event, the speaker can thus conceptualise three different event profiles so the agents have to be able to deal with multiple argument realisation:

- One in which only the puppet 'Jill' is expressed (playing the participant role 'move-inside-1');
- One in which only the object 'house' is expressed (playing the participant role 'move-inside-2');
- One in which both participants are expressed.

The meaning of events is a simple copy of the facts in the memory of the agent, so the meaning of the move-inside-event would simply be conceptualised as:

(6.2)
```
(move-inside event-163190 true)
(move-inside-1 event-163190 jill)
(move-inside-2 event-163190 house-1)
```

For those objects that have to be expressed explicitly according to the event profile, the agent plays a simple discrimination game (Steels, 1996b, 1997a). Suppose that the agents observe a scene in which there are two blocks represented as the following facts:

(6.3)
```
(ball object-1) (blue object-1)
(ball object-2) (green object-2)
```

If the agent wants to talk about the blue ball, it needs a feature or a feature set which discriminates this ball from the green one. Since both objects have the feature 'ball', this cannot be used for discriminating the blue ball from the green one. The colour-feature, however, is discriminating so the speaker would conceptualise the following meaning for the blue ball:

(6.4)
```
(blue object-1)
```

If there are more than one discriminating features (e.g. the features 'girl' and 'jill' in example 6.5 both discriminate the puppet 'Jill' from the puppet 'Jack'), the agent randomly chooses one. The objects are defined in such a way that there is always at least one feature discriminating them from other objects in the same scene.

Suppose that the speaker profiles the move-inside-event such that only the house-object has to be expressed explicitly – roughly meaning something like '(something) moved inside the house / the yellow thing'. Conceptualisation could then possibly yield the following meanings:

(6.5)
```
(move-inside event-163190 true)
(move-inside-1 event-163190 jill)
(move-inside-2 event-163190 house-1)
(yellow house-1)
```

6.1.4 Additional assumptions and scaffolds.

The agents in these simulations are endowed with strong (cognitive) capacities that enable them to communicate linguistically with each other. Listing all aspects of language that are assumed, scaffolded or ignored would take too much space, so I will restrict myself to the most important ones:

- The agents are assumed to be social and cooperative. All agents are equally involved in the formation of their language, so the simulations ignore the possible influence of differences in social status;
- All agents are 'adult' language users. No growth of cognitive capabilities occurs and the models do not take specific child language acquisition restrictions into account. All agents are endowed with the same capacities;
- The agents are assumed to be able to communicate about compositional meanings (i.e. meanings which are related to each other in some way). In the simulations, Fluid Construction Grammar is used as a formalisation of the capacity to combine and manipulate hierarchical symbolic form-meaning mappings. The models further ignore how these symbolic units should be coupled to neurological processing;
- The agents do not have real 'speech'. Instead, all utterances are perfectly transmitted from the speaker to the hearer in the form of strings of words (and markers). The influence of phonological changes on grammaticalisation processes is well understood in theoretical linguistics, but computational models on the formation of phonological and syllabic conventions are scarce and not advanced enough to be used for example to model phonological reduction (although see Steels & Kaplan, 1998b). The phonological development of case markers is therefore ignored in this thesis, but remains a topic of interest for the future;

- The agents also do not care about morphology. There is no meaningful word-internal structure and the capacity of segmentation is assumed. The language-specific segmentation process is scaffolded so the agents can perfectly cut up utterances into words and markers. The markers themselves can thus be seen as adpositions rather than as true markers found in case languages such as German.

Another very important scaffold is the fact that the agents start with a **predefined lexicon, but no grammar**. There are several reasons for giving the agents a lexicon in advance. One reason is methodological: as in any other kind of controlled experiment, this thesis focuses only on the formation of a case grammar and not on the formation of a lexicon. Other experiments using the design stance methodology have already extensively investigated how adaptive lexicons can be formed and shared by large populations of agents (see section 3.3.1). All forces working on the lexicon are therefore completely scaffolded. This also means that the meanings of lexical entries are fixed and known by the agents, so they never have to perform word sense disambiguation either. One of the future steps of the research program would, however, involve the integration of lexical and grammatical development in order to verify whether the current results and conclusions remain valid.

A second reason has to do with a very important assumption that I briefly discussed in section 2.3, which is that the mechanisms underlying the *very first* formation of grammatical constructions are the same ones as those identified in attested cases of present-day grammaticalization processes. These cases (almost) always display a development from more lexical to more grammatical functions:

> Grammaticalization is the gradual drift in all parts of the grammar toward tighter structures, towards less freedom in the use of linguistic expressions at all levels. Specifically, lexical items develop into grammatical items in particular constructions [...]. In addition, constructions become subject to stronger constraints and come to show greater cohesion.
> Haspelmath (1998, p. 318)

So the agents start from a lexicon and build their grammar on top of that. This strategy is however not uncontroversial in the field artificial language evolution. For example, Wray (1998, 2002) argues that modern language evolved through the analysis of a holistic protolanguage (see section 3.1). Similarly, many simulations (especially in Iterated Learning Models, see section 3.1.2) feature the analysis of holistic utterances into smaller linguistic units. Apart from the many arguments against the realism of the 'holistic utterances-first' hypothesis (see De Pauw, 2002, p. 345–348; and Wellens *et al.*, 2008), the most compelling argument for starting from a lexical language is Ockham's razor: since there are no data available from the first language(s), we should first of all investigate what can be explained and learned from applying present and attested processes of grammaticalization rather than starting from a hypothetical holistic language (Hoefler & Smith, 2008). If successful, models on the formation of grammar can also be relevant for modeling language change.

One important observation is that even though the agents start with a lexical language, case markers can be constructed in grammatical languages as well (see section 4.2). Rather than thinking of the initial stage as a 'lexical language' it is more fruitful to think of individual constructions which do not mark event structure grammatically as the seedbed for grammatical markers. The existence of such constructions also shows that **grammaticalization is not a determined process**: speakers of a language can but do not have to decide to tighten their linguistic items towards more grammatical uses.

The lexical entries provide the agents with a language which conforms to what Gil (2008, p. 124) calls an 'Isolated-Monocategorial-Associational Language' (IMA):

1. All the words are morphologically isolating (i.e. they have no internal morphological structure);
2. There are no formal grounds to distinguish syntactic categories such as nouns or verbs. The words are thus syntactically monocategorial. There is only a semantic distinction between words that refer to objects and those that predicate and refer to event types;
3. Utterances are semantically associational: no grammar exists for marking event structure so the hearer has to find out herself how meanings relate to each other.

The lexical entries thus look very much as those presented in section 5.4.1 but this time no semantic or syntactic categories are assumed. The entry for words that refer to event types only contain their event-specific participant roles in their meaning but no potential semantic roles (yet). For example, the semantic pole of the entry of *give* lists a giver, a given and a givee (which have been assigned the more neutral and arbitrary labels give-1, give-2 and give-3). The form-feature in the syntactic pole simply looks for the string "give" and does not give any information yet about the word's potential valents. Both poles feature a J-operator which is used for pulling the form and meaning of *give* into a separate unit. The complete entry of *give* looks as follows:

```
<Lexical entry:  give
((?top-unit
   (meaning (== (give ?event-x true)
                (give-1 ?event-x ?obj-x)
                (give-2 ?event-x ?obj-y)
                (give-2 ?event-x ?obj-z))))
 ((J ?new-unit ?top-unit)
   (referent ?event-x)))
<==>
((?top-unit
   (form (== (string ?new-unit "give"))))
 ((J ?new-unit ?top-unit)))>
```

The entry of the word *jack* looks as follows:

```
<Lexical entry:  jack
((?top-unit
    (meaning (== (jack ?object-1))))
 ((J ?new-unit ?top-unit)
    (referent ?object-1)))
<==>
((?top-unit
    (form (== (string ?new-unit "jack"))))
 ((J ?new-unit ?top-unit)))>
```

In the baseline experiments 20 different scenes were recorded featuring 15 distinct event types including ten macro-events (move-inside, move-outside, hide, give, take, cause-move-on, touch, grasp, fall and walk-to) and five micro-events (borderscreen, visible, move, distance-decreasing and approach). Every micro-event type (except for 'borderscreen') can take the value of true or false which leads to word-pairs such as *visible* versus *invisible*. This makes up for 19 different words that can be used for referring to events. The total number of event-specific participant roles is 30 resulting from three one-place predicates (borderscreen, visible and move), nine two-place predicates (move-inside, move-outside, hide, touch, grasp, fall, walk-to, distance-decreasing and approach) and three three-place predicates (give, take and cause-move-on). The event types and the corresponding words are summarised in Table 6.2.

Next to the words for event types, the agents are given 11 unambiguous words for referring to objects. These words map in a one-to-one relationship with facts about objects in the memories of the agents. The words are: *blue, block, boy, girl, green, ground, house, jack, jill, table* and *yellow*. These words can be used for referring to the seven objects that occur in the twenty recorded scenes (a puppet called 'Jill', a puppet called 'Jack', a blue block, a green block, a table, a yellow house and the ground).

6.2 Baseline experiment 1: no marking

The debate on holistic versus compositional languages (see section 3.1) implicitly assumes a dichotomy between languages that are either holistic or compositional in a grammatical way. However, 'compositional' does not necessarily mean 'grammatical'. There is at least one additional possibility and that is a language which relies heavily on content words rather than on grammatical constructions. In the extreme case this is an entirely lexical language which also forms the first stage in the experiments reported in this thesis. The goal of this first baseline experiment is to **demonstrate that agents can infer the speaker's intended meaning without using grammar but by exploiting the situatedness of the language game**.

Event-types	Participant roles	Truth values	Words
borderscreen	object-1	false	“disappear”
move	move-1	true	“move”
		false	“rest”
visible	visible-1	true	“visible”
		false	“invisible”
approach	approach-1 approach-2	true	“approach”
		false	“no-approach”
distance-decreasing	distance-decreasing-1 distance-decreasing-2	true	“get-closer”
		false	“separate”
fall	fall-1 fall-2	true	“fall”
grasp	grasp-1 grasp-2	true	“grasp”
hide	hide-1 hide-2	true	“hide”
move-inside	move-inside-1 move-inside-2	true	“move-inside”
move-outside	move-outside-1 move-outside-2	true	“move-outside”
touch	touch-1 touch-2	true	“touch”
walk-to	walk-to-1 walk-to-2	true	“walk-to”
cause-move-on	cause-move-on-1 cause-move-on-2 cause-move-on-3	true	“cause-move-on”
give	give-1 give-2 give-3	true	“give”
take	take-1 take-2 take-3	true	“take”

Table 6.2: This table gives an overview of the different event types that occur in the baseline experiments. For each event type, it is specified what participant roles it takes, what truth-values are possible and which word is used for it.

6.2.1 An inferential coding system

Languages across the world vary a lot as to which aspects of the meaning are expressed using grammatical constructions and which are left implicit in the message. Speakers are nevertheless capable of filling in the blanks and reaching communicative success. This is possible because language is an inferential coding system (Sperber & Wilson, 1986) in which the interpreter is assumed to be intelligent enough to infer the correct meaning by using all the possible resources at hand such as the shared context and previous experience. This view is nicely put as follows by Langacker (2000, p. 9):

> It is not the linguistic system *per se* that constructs and understands novel expressions, but rather the language user, who marshalls for this purpose the full panoply of available resources. In addition to linguistic units, these resources include such factors as memory, planning, problem-solving ability, general knowledge, short- and longer-term goals, as well as full aprehension of the physical, social, cultural, and linguistic context.

As a consequence the representation or categorisation system (in this case language) can be much more compact and does not encode the entire meaning. This is different from 'Shannon coding' which is typically used in computer programs where all the information is stored and fixed in the message. The capacity of inferring the intended meaning of the speaker on the basis of partial linguistic input is a **necessary prerequisite** for the formation of grammar in a cognitive-functional framework: without it, innovation and learning would be impossible.

Applied to the topic of this thesis, the first key cognitive ability that the agents need is therefore the capacity to find out how the meanings of the individual words uttered by the speaker should be linked to each other. This is implemented using the same formalisation of meanings as presented in section 5.2 of the previous chapter. I will illustrate this with an example. Suppose that the speaker and hearer both observe a scene in which the puppet 'Jack' walks towards the puppet 'Jill' and that sensory-motor processing yields the following facts in the memory of the agents:

(6.6)
```
(boy object-1) (jack object-1)
(girl object-2) (jill object-2)
(walk-to event-1 true)
(walk-to-1 event-1 object-1)
(walk-to-2 event-1 object-2)
```

The speaker. The speaker first conceptualises a meaning for communicating to the hearer. First the event is profiled and then a discriminating description is found for all the participants that have to be expressed explicitly according to this event profile. Let's assume that the speaker profiles the entire event so both participants need to be expressed. For both puppets, there are two distinctive features in the fact-base so the speaker randomly chooses one for each. This yields the following meaning after conceptualisation:

(6.7)
```
(jack object-1)
(girl object-2)
(walk-to event-1 true)
(walk-to-1 event-1 object-1)
(walk-to-2 event-1 object-2)
```

Next, the speaker starts a production task which runs entirely as explained in the previous chapter. Since the speaker only has a lexical language, only the lexical information of the entries for *walk-to, jack* and *girl* is added. This results in the following node in the speaker's reaction network:

```
<Node-3:  coupled-feature-structure
((top-unit
   (sem-subunits (jack-unit walk-to-unit girl-unit)))
 (jack-unit
   (meaning ((jack object-1)))
   (referent object-1))
 (walk-to-unit
   (referent event-1))
   (meaning ((walk-to event-1 true)
             (walk-to-1 event-1 object-1)
             (walk-to-2 event-1 object-2))))
 (girl-unit
   (meaning ((girl object-2)))
   (referent object-2)))
<==>
((top-unit
   (syn-subunits (jack-unit walk-to-unit girl-unit)))
 (jack-unit
   (form ((string jack-unit ”jack”))))
 (walk-to-unit
   (form ((string walk-to-unit "walk-to"))))
 (girl-unit
   (form ((string girl-unit ”girl”)))))>
```

In the original experiments, each unit also had a 'goal'-feature in the semantic pole with the values 'assert' for the top-unit, 'reference' for the units of the participants, and 'predicate' for the units of the event types. In the syntactic pole, there was also a 'scope'-feature (e.g. with the value 'utterance' for the top-unit). Since they play no role in the simulations, I left them out in the replicating experiments. Since the speaker has unified and merged all the possible linguistic items, we get the following utterance (word order is completely random because it was not specified in the coupled feature structure):

(6.8) "walk-to jack girl"

The hearer. The hearer observes the speaker's utterance and starts parsing it. This involves segmenting the utterance into words and then unifying and merging all possible linguistic items with the current node in the reaction network. The hearer, too, only knows lexical words so only lexical information is added to the coupled feature structure. This results in the following node:

```
<Node-3:  coupled-feature-structure
((top-unit
   (sem-subunits (jack-unit walk-to-unit girl-unit)))
 (jack-unit
   (meaning ((jack ?object-a)))
   (referent ?object-a))
 (walk-to-unit
   (referent ?event-x))
   (meaning ((walk-to ?event-x true)
             (walk-to-1 ?event-x ?object-x)
             (walk-to-2 ?event-x ?object-y))))
 (girl-unit
   (meaning ((girl ?object-b)))
   (referent ?object-b)))
<==>
((top-unit
   (syn-subunits (jack-unit walk-to-unit girl-unit))
   (form ((meets jack-unit walk-to-unit)
          (meets walk-to-unit girl-unit))))
 (jack-unit
   (form ((string jack-unit ”jack”))))
 (walk-to-unit
   (form ((string walk-to-unit "walk-to"))))
 (girl-unit
   (form ((string girl-unit ”girl”)))))>
```

The hearer can now extract the following meaning from the semantic pole of this coupled feature structure:

(6.9)
```
((jack ?object-a) (girl ?object-b) (walk-to ?event-x true)
 (walk-to-1 ?event-x ?object-x)
 (walk-to-2 ?event-x ?object-y))
```

Note that the hearer does not know from this meaning which participant played which role in the event. Both Jack and Jill could be the participant which is moving towards the other, or perhaps even another (implicit) participant plays a role. The hearer can **interpret** the parsed meaning by unifying it with the facts in the memory, which yields the following set of bindings:

(6.10)
```
((?event-x .  event-1) (?object-x .  object-1)
 (?object-y .  object-2) (?object-a .  object-1)
 (?object-b .  object-2))
```

Since unification was successful and returned only one hypothesis-set, the hearer can now infer that the variables '?object-x' and '?object-a' are equal because they both refer to the same object ('object-1'). The hearer also infers that the variables '?object-y' and '?object-b' are equal because they both refer to 'object-2'. The hearer thus successfully infers that the puppet Jack played the participant role 'walk-to-1' and that the puppet Jill played the participant role 'walk-to-2'.

Communicative success. In the above example, the parsed meaning is unambiguously compatible with the hearer's world model so the hearer signals agreement to the speaker. This means that the language game was successful. The game fails if the parsed meaning would not match the hearer's world model or if the hearer still has multiple hypotheses left after interpretation (for example when the context consists of several similar events). In this case the hearer would signal disagreement.

In this first baseline experiment, success in the game does not influence the linguistic behaviour of the agents since the lexicon is given and assumed to be fixed. The point here is rather to demonstrate that the agents can reach success in communication even though they have no grammar yet.

6.2.2 Results and discussion

The above experimental set-up was tested in **a population of two and a population of ten agents** engaging in peer-to-peer description games without a cross-generational population turnover. In each game, the agents share a context of five events from the same scene. Two measures were used: **communicative success** and **cognitive effort** (see Appendix A for a description of all measures).

Results. The results of baseline experiment 1 are illustrated in Figure 6.3. The graph displays communicative success and cognitive effort for 10 series of 500 language games. Since the language of the agents is given and static, and since the task difficulty never changes, the results show a constant behaviour over time. Experiments using a population of ten agents yielded the same results for the same reasons.

The results indicate that the agents can indeed reach a fair amount of communicative success without using grammar: in about 70% of the games, the hearer is capable of unambiguously inferring the intended meaning from the context. For this, the hearer needs an average cognitive effort of 0,6 during interpretation. Cognitive effort is fairly high since all failed games are counted as requiring maximum cognitive effort of 1. If the results of the failed games are ignored, cognitive effort drops to 0,5 on average. The simulations were run using the skewed frequency of event types.

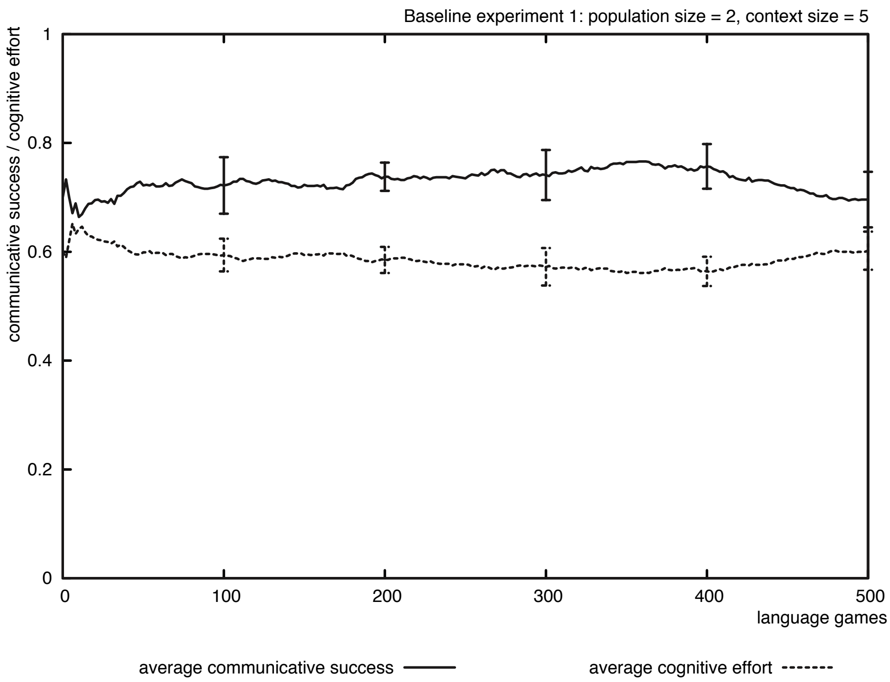


Figure 6.3: This graph shows the average cognitive effort and communicative success in baseline experiment 1 for 10 series of 500 language games in a population of two agents and a context size of five events. Success is reached in about 70% of the games. Cognitive effort during interpretation amounts to 0,6 on average.

Discussion. The results of baseline experiment 1 demonstrate that the agents can still reach communicative success if they use their world model for inferring the intended meaning. The proposed machinery thus works but only under certain conditions. For one thing, the hearer needs to have witnessed the scene in order to make the correct inferences. Second, the context cannot be too ambiguous otherwise interpretation can involve multiple hypotheses. As can be read from the average communicative success measure, this happens in 30% of the cases. Failed games typically occur when the scene contains at least two event tokens which have the same event type but which involve different participants (e.g. both puppets are moving).

Improving communicative success in the failed games could be achieved in many ways: agents can be given more complex dialogue strategies, the speaker can use pointing, the hearer can be more bold in making assumptions about the speaker's intention or ask for additional feedback, etc. These strategies would however involve more negotiation and do not reduce the cognitive load during interpretation for the hearer. Additional marking, however, would be a solution which could resolve ambiguity and reduce cognitive effort during interpretation at the same time. This solution will be tested in the next baseline experiment.

6.2.3 Conclusions

In the first baseline experiment, a population of two and a population of 10 agents engaged in a series of description games. The agents were endowed with the capacity to unify parsed meanings with their world models from which they can correctly infer the speaker's intended meaning (a) if the description is compatible with the world model, and (b) if unification does not yield multiple hypotheses. The results showed that if these two requirements are met, a lexical language suffices for reaching communicative success. The drawback is however that agents always need to witness the same scene (i.e. displacement becomes very hard and impossible in some cases) and the agents need a lot of cognitive effort for inferring the intended meaning.

6.3 Baseline experiment 2: specific marking

Baseline experiment 1 showed how agents can still reach communicative success without using grammar. In the second baseline experiment, agents can exploit this ability to autonomously detect whether it might be useful to make changes to their linguistic inventories in order to optimise communication. The hypothesis investigated here can be formulated as follows: **ambiguity or too much cognitive effort during parsing and interpretation can be a trigger for the invention of functional markers for optimising communicative success**.

6.3.1 Speaker-based innovation

In baseline experiment 1 the hearer is faced with the cognitive load of figuring out who's doing what in events during each interaction. Moreover, if the context is complex enough it may not be clear which event the speaker was referring to. In this experiment, the agents are therefore endowed with a second key cognitive ability which involves the **innovation and expansion** of their language through event-specific markers. This ability comprises three subparts:

1. Expanding the agents architecture with a 're-entrant' mapping for detecting opportunities for optimisation and learning innovations;
2. Endowing the agents with the capacity of inventing a marker and the corresponding constructions;
3. Endowing the agents with a consolidation mechanism which allows them to converge on a shared set of markers.

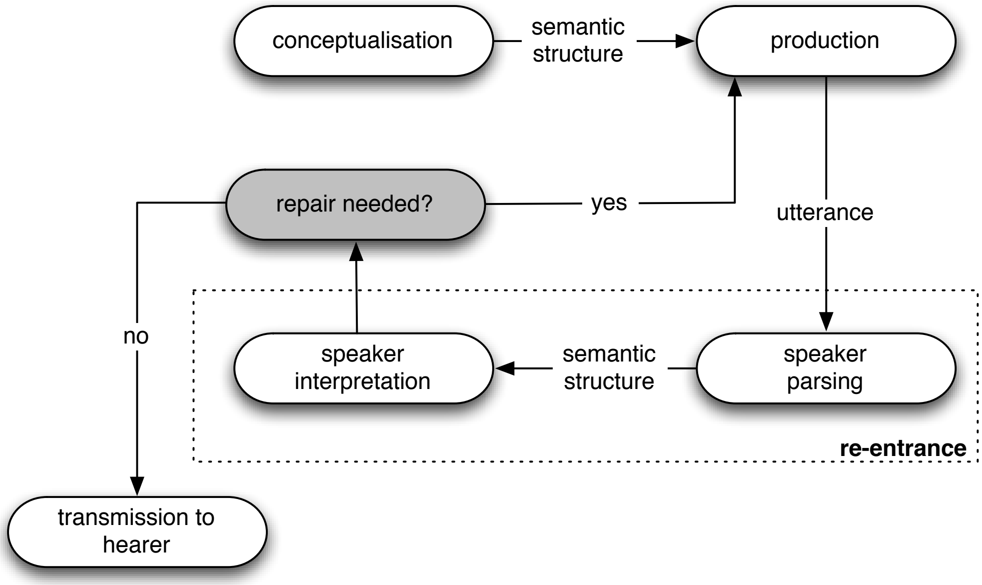


Figure 6.4: Before transmitting an utterance to the hearer, the speaker 're-enters' her own utterance in her language system and uses herself as a model of the hearer. In this way the speaker estimates whether there might be problems or too much complexity for the hearer during parsing. If so, the speaker tries to repair the problem.

re-entrance. In baseline experiment 1, the agents could confidently use their lexical items to communicate with each other since the lexicon in this model is fixed, unambiguous, and shared by all the agents. However, should this scaffold be taken away, the agents would have to worry about whether the words they use are also known and understood by the other agents. They would thus somehow have to be capable of predicting the parsing and interpretation behaviour of the hearer in order to increase the chances of reaching communicative success. This can be achieved through 're-entrance' (Steels, 2003b, also called the 'obverter' strategy by Smith, 2003a).

Re-entrance can be thought of as self-monitoring in which the speaker does not directly transmit her utterance to the hearer, but first 're-enters' the utterance into her linguistic system and parses and interprets the utterance herself as if she was the hearer. By taking herself as a model to simulate the linguistic behaviour of the hearer, the speaker can detect whether there might be problems or difficulties during parsing and interpretation. If so, the speaker will try to solve this problem. This strategy is illustrated in Figure 6.4. Similarly, the hearer can also use a re-entrant mapping for simulating the behaviour of the speaker. Technically speaking, achieving re-entrance is not so difficult since the agents in these experiments can act both as a speaker and as a hearer.

Re-entrance is thus a crucial strategy in inferential coding systems: if the speaker wants to solve a problem through innovation, she needs to have an educated guess about the hearer's knowledge and which aspects of the common ground can be exploited for getting the message across. The hearer has to perform the same kind of reasoning for guessing the speaker's intentions. Human language users obviously adapt their linguistic behaviours to their speech partners (e.g. when speaking to children or second language learners). Given the fact that all agents in the experiments are each other's peers, the best model they can have of other agents is themselves.

Innovation. It is unavoidable that language users come across situations in which the speaker does not know an adequate and well-entrenched convention for expressing a certain meaning, especially in the extreme case where there is no grammar at all. In this experiment, the speaker will invent a specific marker for explicitly expressing a particular participant role if there are possible ambiguities in the context or if the hearer needs to do more inference than is desirable. This is implemented through **diagnostics and repair strategies** which run in this experiment along the following algorithm:

1. **Diagnostic 1:** Re-enter the utterance into the linguistic system and start a parsing and interpretation task.
    a. If interpretation returns a failure or multiple hypotheses, report a problem;
    b. If there is one possible hypothesis which contains at least one unexpressed variable equality, report a problem;
    c. If there is no inference needed and if there is only one hypothesis, transmit the utterance to the hearer.
2. **Repair strategy 1:** If a problem of ambiguity or unresolved variable equalities has been reported, trigger repair strategy 1.
    a. If there is only one unexpressed variable equality, invent a new marker for it and start a new production task;
    b. If there are more than one unexpressed variable equalities left (i.e. the repair is too difficult), ignore the problem and transmit the utterance to the hearer.

I will now illustrate this algorithm with a concrete example. Suppose that the speaker and hearer both observe the same scene as the one used in section 6.2 in which Jack was walking to Jill; and that the sensory-motor processing delivers the same facts as in example 6.5. This time the speaker only profiles the part of the event involving Jack and conceptualises the following meaning:

(6.11)
```
(jack object-1)
(walk-to event-1 true)
(walk-to-1 event-1 object-1)
(walk-to-2 event-1 object-2)
```

Since the speaker has no grammar yet, only the lexical entries of *Jack* and *walk-to* are unified and merged with the coupled feature structure which results in the following, randomly ordered utterance:

(6.12) "jack walk-to"

Instead of directly transmitting the utterance to the hearer, the speaker first re-enters the utterance into her own linguistic system and parses the following meaning:

(6.13)
```
(jack ?object-a)
(walk-to ?event-1 true)
(walk-to-1 ?event-x object-x)
(walk-to-2 ?event-x object-y)
```

Interpreting this meaning by unifying it with the speaker's world model yields the following set of bindings:

(6.14)
```
((?event-x . event-1) (?object-x . object-1)
 (?object-y . object-2) (?object-a . object-1))
```

Unification is successful and returns only one hypothesis so the speaker does not detect ambiguity. However, there is a variable equality left between the variable '?object-x' and '?object-a': both refer to 'object-1'. **Diagnostic 1 will therefore report a problem which triggers repair strategy 1**.

The repair strategy assesses the difficulty of the problem: if there are more than one unexpressed variable equalities left, the problem is classified as 'too difficult to solve' and then the utterance is transmitted anyway. Here, however, there is only one variable equality so the speaker will invent a new marker for it. This marker is specific to the walk-to-1 role and can thus almost be seen as a lexical item itself. The speaker then invents a **verb-specific construction** in which the new marker (let's say *-bo*) binds the referent of the walk-to-1-role to the referent of the argument that fills this role by using the same variable '?object-x'. The syntactic pole states that this argument plays 'syn-role-1' which is nothing more than a direct mapping of the walk-to-1-role.

```
<Construction: construction-1
((?unit-1
   (meaning (== (walk-to ?event-x ?truth)
                (walk-to-1 ?event-x ?object-x)
                (walk-to-2 ?event-x ?object-y))))
 (?unit-2
   (referent ?object-x)))
<==>
((?unit-2
   (syn-role syn-role-1)))>
```

The morphological rule states that the marker immediately follows the argument which plays the walk-to-1-role. As explained in the previous chapter, the morphological rule will create a new marker-unit and make it a sub-unit of the argument-unit. Both the verb-specific construction and the morph-rule look slightly different from the original proposals by Steels (2002b) due to changes in the grammar formalism, but they have the same performance.

```
<Morph-rule: -bo
((?top-unit
   (syn-subunits (== ?unit-1)))
 (?unit-1
   (syn-role syn-role-1)))
<==>
((?top-unit
   (form (== (string ?marker-unit "-bo")
             (meets ?unit-1 ?marker-unit))))
 ((J ?marker-unit ?top-unit))
 ((J ?unit-1 ?top-unit (?marker-unit))))>
```

The speaker now starts a new production task for the same meaning. In the initial node in the reaction network, the whole meaning is still grouped together in one unit. The speaker then unifies and merges the lexical entries of *jack* and *walk-to* with the initial node which leads to a separate unit for each word. The new coupled feature structure looks as follows:

```
<Node-2: coupled-feature-structure
((top-unit
   (sem-subunits (jack-unit walk-to-unit)))
 (jack-unit
   (meaning ((jack object-1)))
   (referent object-1))
 (walk-to-unit
   (referent event-1))
   (meaning ((walk-to event-1 true)
             (walk-to-1 event-1 object-1)
             (walk-to-2 event-1 object-2))))
<==>
((top-unit
   (syn-subunits (jack-unit walk-to-unit)))
 (jack-unit
   (form ((string jack-unit ”jack”))))
 (walk-to-unit
   (form ((string walk-to-unit "walk-to")))))>
```

Before the repair, this would be the final node in the reaction network. This time, however, the speaker has the newly made construction at her disposal. Since this is a production task, the semantic pole of the construction needs to unify with the semantic pole of node-2. This is successful: the construction needs any unit containing the meaning of a walk-to-event and another unit of which the referent is the same one as the referent of the walk-to-1-role ('object-1'). The syntactic pole of the construction then simply merges the feature-value pair '(syn-role syn-role-1)' to the argument-unit. The construction thus licenses the following node in the reaction network:

```
<Node-3:  coupled-feature-structure
((top-unit
    (sem-subunits (jack-unit walk-to-unit)))
 (jack-unit
    (meaning ((jack object-1)))
    (referent object-1))
 (walk-to-unit
    (referent event-1))
    (meaning ((walk-to event-1 true)
              (walk-to-1 event-1 object-1)
              (walk-to-2 event-1 object-2))))
<==>
((top-unit
    (syn-subunits (jack-unit walk-to-unit)))
 (jack-unit
    (syn-role syn-role-1)
    (form ((string jack-unit ”jack”))))
 (walk-to-unit
    (form ((string walk-to-unit "walk-to")))))>
```

Next, the speaker can unify and merge the new morph-rule with node-3. The left-pole of the morph-rule (which is syntactic, see section 5.4.2) looks for any unit containing the feature-value pair '(syn-role syn-role-1)' which is indeed present in the syntactic pole of node-3. Next, the right-pole of the morph-rule is merged with the right-pole of node-3:

```
<Node-4:  coupled-feature-structure
((top-unit
    (sem-subunits (jack-unit walk-to-unit)))
 (jack-unit
    (meaning ((jack object-1)))
    (referent object-1))
 (walk-to-unit
    (referent event-1))
```

```
   (meaning ((walk-to event-1 true)
            (walk-to-1 event-1 object-1)
            (walk-to-2 event-1 object-2))))
<==>
((top-unit
   (syn-subunits (jack-unit walk-to-unit)))
 (jack-unit
   (syn-subunits (bo-unit))
   (syn-role syn-role-1)
   (form ((string jack-unit ″jack″))))
 (bo-unit
   (form ((string bo-unit ″-bo″) (meets jack-unit bo-unit))))
 (walk-to-unit
   (form ((string walk-to-unit "walk-to")))))>
```

The speaker has no other items that can be unified and merged so node-4 is the final node in the reaction network. The speaker then renders the form-features of the syntactic pole into an utterance. The ordering between the words *jack* and *walk-to* are still random, but the coupled feature structure specifies that the marker *-bo* immediately follows *jack*. This results in the following utterance:

(6.15) “jack -bo walk-to”

The speaker now re-enters this utterance again into her linguistic system to check whether the innovation has the intended effect. I will not repeat the entire trace of parsing here since this is completely analogous to the example given in Chapter 5. Parsing the utterance yields the following meaning:

(6.16)
```
(jack ?object-x)
(walk-to ?event-x true)
(walk-to-1 ?event-x object-x)
(walk-to-2 ?event-x object-y)
```

Note that this time, the meaning-predicates ‘jack’ and ‘walk-to-1’ share the same variable ‘?object-x’. Interpretation then returns the following set of bindings:

(6.17)
```
((?event-x .  event-1) (?object-x .  object-1)
 (?object-y .  object-2))
```

As can be seen in the set of bindings, there are no unexpressed variable equalities left so no additional inferences are needed. The speaker is thus satisfied with the utterance and transmits it to the hearer.

Learning. Learning a new marker is very similar to inventing one and is achieved through the same cognitive mechanisms. The hearer first observes the utterance and then parses it. If there are unknown strings, such as the marker *-bo* which was just invented by the speaker, the hearer will ignore it and try to parse as much as possible. Then the hearer tries to interpret the parsed meaning. If there are unexpressed variable equalities left, the same diagnostic as was used by the speaker will report a problem. The repair strategy then tries to find out whether the utterance contains elements which could carry this particular meaning or function.

1. **Hearer diagnostic 1:** Parse the utterance and interpret its meaning.

   a. If interpretation returns a failure or multiple hypotheses, report a problem;

   b. If there is one possible hypothesis which contains at least one unexpressed variable equality, report a problem;

   c. If there is no inference needed and there is only one hypothesis, signal agreement to the speaker.

2. **Hearer repair strategy 1:** If a problem of ambiguity or unresolved variable equalities has been reported, trigger repair strategy 1.

   a. If there is only one unexpressed variable equality, check whether there was one unknown string in the utterance. If so, add a new verb-specific construction to the inventory which records the unknown string as a marker for the variable equality. If not, ignore the problem and signal agreement or disagreement to the speaker depending on success of the game;

   b. If there are more than one unexpressed variable equalities left or if there were multiple unknown strings, ignore the problem. Transmit success to the hearer if inference is nevertheless possible.

   c. If interpretation fails or leads to multiple hypotheses, ignore the problem and signal disagreement.

Consolidation. In the original two-agent simulations variety never occurs since the agents only observe each other's inventions except for the extremely rare cases in which the learning task was too difficult and the learner later on invents a different solution for the same problem. So consolidation is fairly trivial and means just storing the newly created or learned items in the linguistic inventory.

However, as soon as we scale up the experiments to multi-agent populations involving at least three agents, a pool of synchronic variation naturally arises since the agents can independently come up with different innovations for the same problems (as I discussed in section 3.1.3). The agents therefore need to have an alignment strategy which enables them to deal with the variety and to converge on a shared set of preferred markers.

In section 3.3.2 I argued that the experiments on grammar first of all try to move all the previous work on lexicon formation onto the domain of grammar, so this experiment starts with the a similar alignment strategy as was suggested in prior work. This strategy goes as follows: each construction has its own **confidence score** between 0 and 1. The higher the score, the more confident the agent is that the item is a conventionalised unit in the population. Based on the game's success and based on the speaker's behaviour, the hearer will update the scores in the inventory as follows:

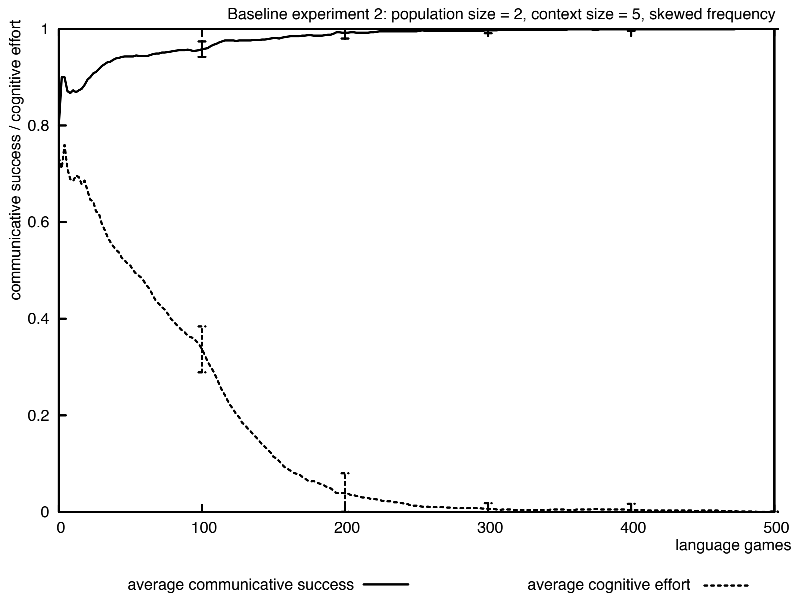
- In case of success, increase the score of the applied construction(s) by 0.1 and decrease the scores of all its competitors through **lateral inhibition** by 0.1. Competitors of a construction are constructions which either have the same semantic pole but a different form (competing synonyms), or the same form but different semantics (competing homonyms);
- If the game was a failure, do nothing.

The fact that only the hearer performs score updating captures the intuition that agents first of all want to **conform to** the behaviours of other agents in the population rather than imposing their own preferences. A mathematical model by De Vylder (2007) also shows that this strategy results in smoother convergence dynamics. In case of game failure, neither the speaker nor the hearer updates any scores. The reason for this is that the description game does not offer enough explicit feedback for the agents to find out whether there might be parts in the processing chain which were harmful for communication.

6.3.2 Results and discussion

The above diagnostics and repair strategies have been implemented and tested in three different simulations. The first series (set-up 2a) features a population of two agents and replicates the results obtained by Steels (2002b, 2004c). The second experiment (set-up 2b) involves a population of ten agents in which the consolidation strategies become necessary for convergence. Both experiments feature the skewed frequency of event types mentioned in section 6.1.3. A third set-up (set-up 2c) also features 10 agents, but this time the skewed frequency was replaced by an equal frequency of event types in order to study the convergence and competition dynamics more easily. All the results were obtained after ten series of language games for each simulation.

Results of set-up 2a. The results obtained from the replication experiment confirm the results of the original case experiment. The top graph in Figure 6.5 shows that the average cognitive effort needed by the speaker rapidly drops to zero if the agents start inventing and using specific markers to indicate relations between events and their participants. With the markers the agents are also capable of overcoming ambiguity in the context since communicative success rises to 100%. However, there is a price to pay for this optimisation which is shown in the bottom graph: for each participant role, the agents have to learn and store a specific marker in the inventory. In this two-agent simulation, no variation occurs so agreeing on a set of 30 markers is fairly trivial.



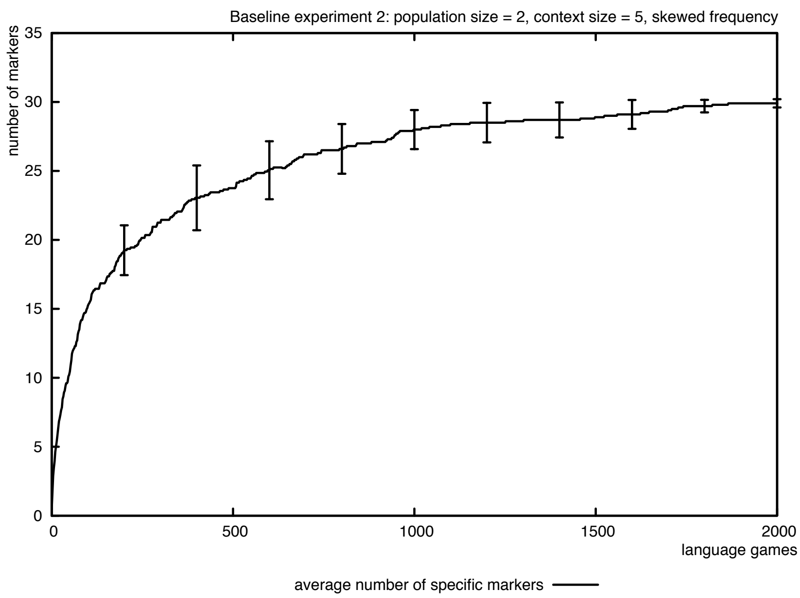


Figure 6.5: The top graph shows that the agents rapidly reach 100% communicative success in the two-agent set-up. The agents also succeed in reducing the cognitive effort to zero. The bottom graph shows that they need to learn and store 30 specific markers in their inventory to do so.

Results of set-up 2b. If the population size is increased, the agents need much more time to learn all the variations floating in the population and to converge on a shared set of 30 markers. As Figure 6.6 shows, the agents need 80.000 language games in order to agree on this set. This is still fairly rapid: it means that each agent needs to play an average of 8.000 games in order to conform to the language of its peers.

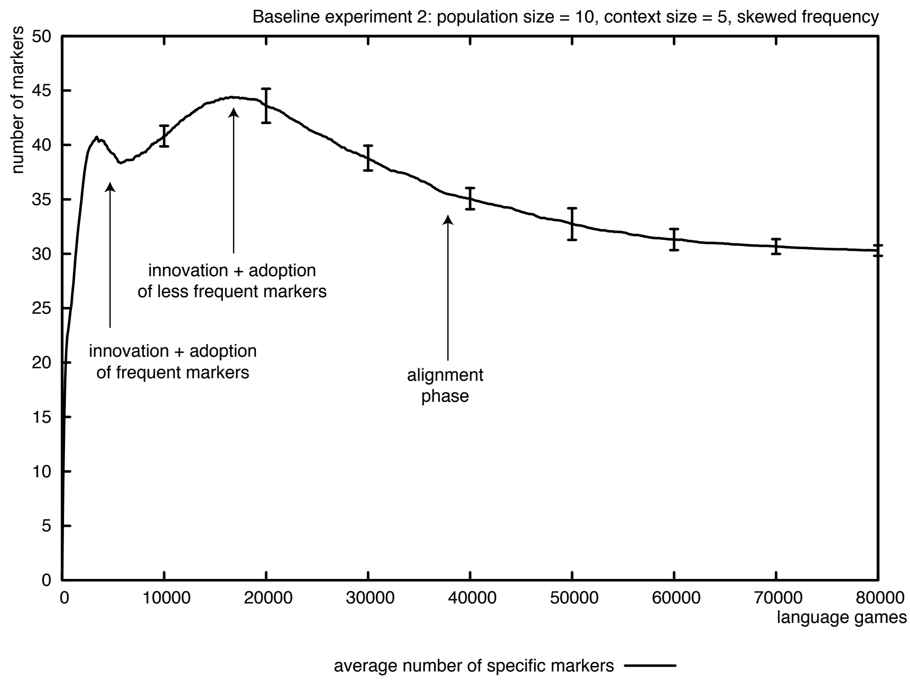


Figure 6.6: This graph shows the number of specific markers in baseline experiment 2b for 10 series of 80.000 language games in a population of 10 agents and a context size of five events. The agents start innovating and learning markers rapidly during the first 3.000 games after which a short period of alignment seems to kick in. The number of markers then rises again between game 7.000 and game 20.000. This is due to the skewed frequency of the event-tyes: events that potentially take three participants are very rare in the data and markers for them are only now being acquired by all agents. Finally, a long alignment period starts which also takes much more time for the less-frequent event types.

The graph first shows a steep rise to an average of 40 markers in the beginning after which an alignment phase seems to start. However, the number of markers starts to climb up again after 7.000 games and reaches a height of about 45 markers by the time 20.000 games have been played. Then there is a long and gradual slope towards convergence at 80.000 games. The two peaks in the graph are the result of the skewed frequency of event type occurrences: markers for three-participant events are very rare and are therefore constructed, learned and propagated later than the frequent markers. Even so, the agents still reach agreement and communicative success while reducing the cognitive effort needed if they are given enough time.

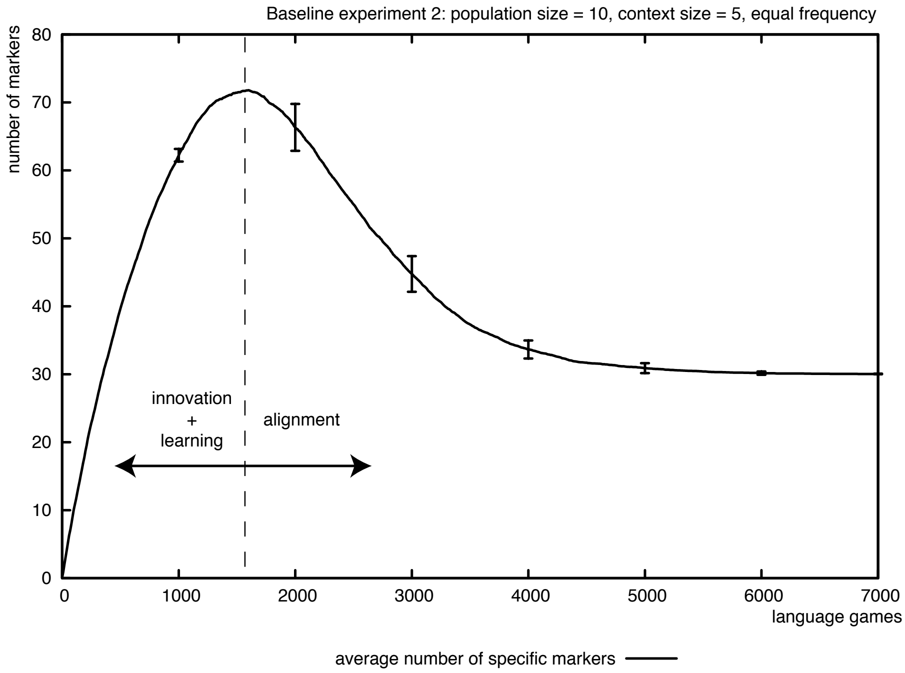


Figure 6.7: This graph shows the number of specific markers in baseline experiment 2c for 10 series of 7.000 language games in a population of 10 agents and a context size of five events. This case, all event types occur with the same frequency so we can see the convergence dynamics in the population more clearly: agents invent and learn new markers during the first 1.500 language games. The agents then need another 4.500 language games in order to align their linguistic inventories with each other.

Results of set-up 2c. In the third set-up all the event types occur with the same frequency so we can better study the convergence dynamics without other influences. Figure 6.7 shows that the agents indeed need significantly less time than in the second set-up: between six and seven thousand language games. On average this means about 600–700 games per agent, which comes close to the 500 games needed by the agents in two-agent simulations. The convergence task here is comparable in difficulty to a multiple word naming game involving 30 objects (see Van Looveren, 2005). The graph shows that the agents keep innovating and learning new markers during the first 1.500 language games after which they rapidly converge on a shared set of 30 markers.

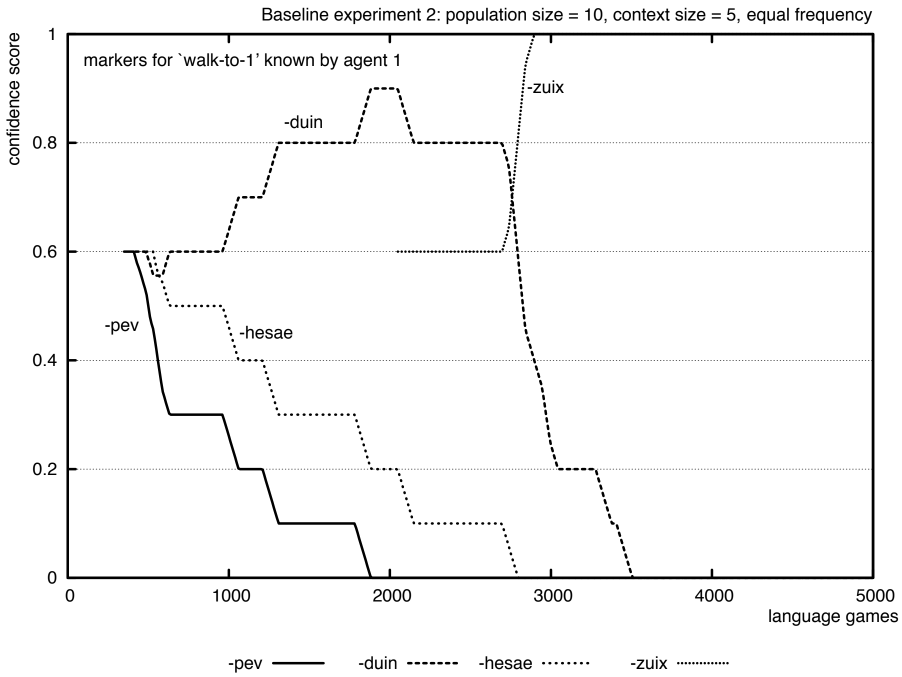


Figure 6.8: This graph shows a snapshot of the competition between forms for marking the participant role 'walk-to-1' within agent 1. Agent 1 learns the marker *-pev* at first, but soon also observes *-duin* and *-hesae*. The marker *-duin* seems to win the competition and even reaches a confidence score of 0.9 after 2.000 language games. However, the agent then learns another marker *-zuix* which is apparently shared by a lot of other agents in the population: at game 3.000 it has already pushed *-duin* down and it reaches 1.0 confidence score.

The peak of 70 markers – as opposed to the peak of 45 with the skewed frequency – is normal since there are more competing markers floating in the population at the same time.

Figure 6.8 gives a snapshot of one agent's knowledge of markers for the participant role 'walk-to-1'. The agent learns three markers at about the same time: *-pev*, *-duin*, and *-hesae*. The marker *-duin* seems to be the winning marker and reaches a confidence score of 0.9 by the time the agents have played 2.000 language games. However, the agent then learns a new marker *-zuix* which seems to be quite successful in the population: its confidence score rapidly increases to the maximum while *-duin* goes downhill very fast because of lateral inhibition. In the end, *-zuix* wins the competition.

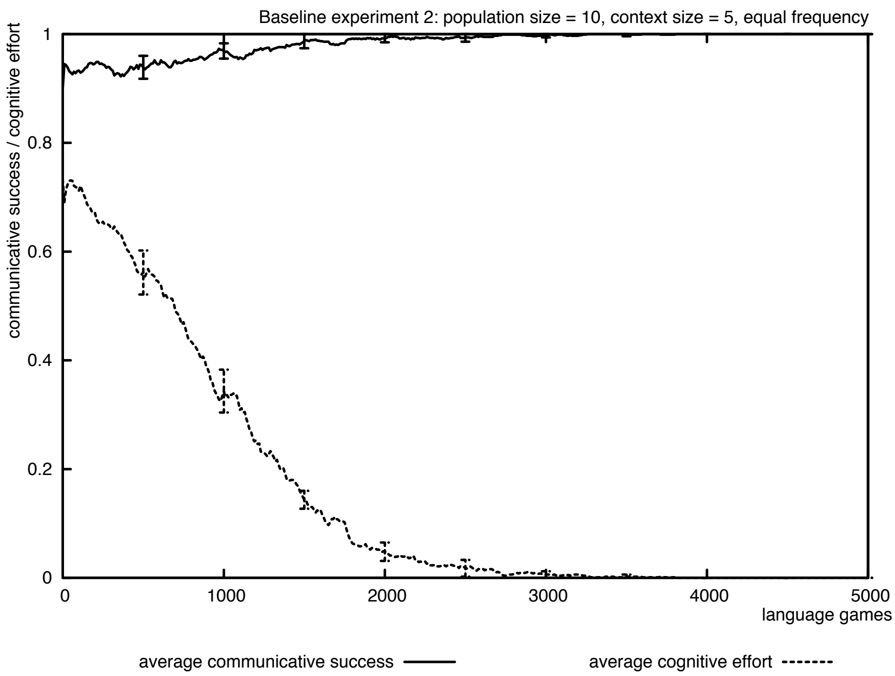


Figure 6.9: In the multi-agent populations too the agents succeed in reaching communicative success as well as reducing the cognitive effort needed for interpretation.

Finally, Figure 6.9 shows the average communicative success and cognitive effort again. The results show that the agents in the multi-agent simulations also rapidly reach 100% communicative success by using markers. Cognitive effort also goes down until no inferences need to be made anymore.

Discussion. The results of baseline experiment 2 clearly illustrate how agents can use locally available information to assess their own linguistic interactions and couple this assessment to repair and consolidation strategies in order to improve communication. In this case, the agents mainly acted to reduce the semantic complexity of interpretation for the hearer: if additional markers are introduced for explicitly indicating the relations between events and their participants, parsing leads the agents immediately to the desired bindings in interpretation. Without the markers, additional inferences would still be needed.

However, a price has to be paid for reducing the effort of interpretation and that is an increased inventory size. During the language games, the agents learn 70 markers on average and retain 30 of them, each one specific to a particular participant role. The inventory size is in fact not the main problem here but a side-effect of a bigger issue: since the markers are restricted to one function only, the agents (given their present capabilities) cannot use them to go beyond the data of known events. Hence there is no generalisation. Inventing new markers for each participant role may be an efficient strategy in a small and fixed world, but becomes problematic once agents have to adapt to new meanings and communicative challenges.

Coupling these results back to natural languages, there are thus two clear qualitative differences: (a) markers in natural languages *do* generalise and become polysemous; and (b) they are not randomly invented but recruited from existing lexical items (see sections 4.2.3 and 4.2.4). Both issues will have to be addressed in other experiments.

6.3.3 Conclusions

In this section I reported three simulations of baseline experiment 2. The first set-up replicated the original case grammar experiments by Steels (2002b, 2004c) and confirmed their results. Next, I reported two multi-agent simulations featuring the same set-up but involving either skewed or equal frequency of event types. In both simulations, a consolidation strategy was used similar to that of prior work on the formation of adaptive lexicons. This strategy enabled the agents to agree on a shared set of 30 specific case markers. Since the simulations involving equal frequency allowed for a clearer understanding of the convergence dynamics, I will keep on using this frequency in subsequent simulations. Once these steps are fully understood, frequency effects on language formation and change can be taken into account again.

The main point of baseline experiment 2 was to demonstrate that additional grammatical marking may be useful for reducing the cognitive effort needed during interpretation and for increasing communicative success. To enable the agents to do this, they used the same mechanisms for inferring meanings from the context to autonomously diagnose problems or opportunities for optimisation during their local language games. The diagnosis was then coupled to a repair strategy which enabled the agents to invent or learn a new case marker. Other self-assessment criteria, such as communicative success and processing information, were fed back into the consolidation strategy enabling the agents to self-organise their linguistic conventions without central control or external guidance by the experimenter.

A new problem arose, however, which is the fact that the agents cannot generalise their markers to go beyond the data. This has consequences on inventory size but may also lead to a reduced communicative success in future interactions if the agents want to express new or more complex meanings. The next experiment investigates how analogical reasoning could be called in to solve this problem.

6.4 Baseline experiment 3: semantic roles

The results of baseline experiment 2 indicate that the agents can reduce the problem of ambiguity and cognitive effort if they make use of additional marking. However, the proposed innovation strategy requires new markers to be invented all the time so (a) there is no generalisation beyond the known data, and (b) this may lead to an explosion of the inventory size in the long run. Baseline 3 investigates how the same principles of diagnosis and repair in order to reach communicative success can lead to generalisation. The hypothesis is that **analogical reasoning over event structures can account for an increased generalisation and productivity of case markers**.

6.4.1 Generalisation as a side-effect

When repairing a problem of unexpressed variable equalties in the previous baseline experiment, the speaker assessed the repair to be too difficult to learn if the context was too ambiguous. In a more complex simulation, however, one can imagine that ambiguity is rather the rule than the exception so the speaker needs to innovate in a more clever way to give the hearer additional clues about what she meant. One such strategy is to reuse existing items as much as possible in semantically related or analogous situations. In baseline experiment 3, the speaker can reuse the existing markers in new situations by performing analogical reasoning over event structures. The analogy algorithm comprises the following steps:

1. Given a target participant role$_i$, find a source role$_j$ for which a case marker already exists;
2. Elaborate the mapping between the two:
   a. Take the target event structure in which participant role$_i$ occurs (provided by sensory-motor processing);
   b. Take the source event structure of the event that was used to create source role$_j$;
   c. Select from the source event structure all the facts and micro-events involving the filler of source role$_j$ and retrieve the corresponding facts and micro-events of the target event structure.
3. Keep the mapping if it is good. A good mapping means that:
   a. the filler of source role$_j$ always maps onto the same object in the corresponding facts and micro-events;
   b. the corresponding object fills the target role$_i$ in in the target structure.
4. If there are multiple analogies possible, choose the best one (based on entrenchment and category size);
5. Build the necessary constructions and make the necessary changes to existing items.

Step 3b in the algorithm ensures that an analogical role is also discriminating enough to distinguish the target role from other possible participant roles in the same or other events. By reusing existing items in novel but similar situations, the speaker reduces the hypothesis space for the hearer and facilitates the abduction process. The hearer can retrieve the analogy using the same algorithm if she also knows the other marker. In this strategy, the generalisation of existing linguistic items is not a goal in itself but rather a side-effect of optimising communicative success.

The target and the source. I will illustrate the analogy algorithm of this experiment through an example. Suppose that the speaker wants to construct a marker for the participant role 'walk-to-2' of the following walk-to-event:

(6.18) `(walk-to event-100 true) (walk-to-1 event-100 jack)`
`(walk-to-2 event-100 jill)`

I will from now on refer to 'walk-to-2' as the *target role* and to 'jill' as the *target filler*. Instead of inventing a new marker immediately, the speaker will first check whether she already knows a marker which is analogous and hence could be reused. Suppose that the speaker already knows the marker *-mi* for the participant role 'move-inside-1'. I will from now on refer to this participant role as the *source role* and its filler as the *source filler*. The speaker has stored the original event which she used for creating the marker and retrieves it from her memory:

(6.19) `(move-inside event-163190 true)`
`(move-inside-1 event-163190 jill)`
`(move-inside-2 event-163190 house-1)`

Elaborate the mapping between the two. In order to elaborate the mapping between the two events, the complete event structure is taken. The target event (walk-to) consists of four micro-events: one participant is moving and approaching another participant which stands still until the two participants touch each other:

- `(move event-165641 true) (move-1 event-165641 jack)`
- `(move event-165419 false) (move-1 event-165419 jill)`
- `(approach event-165486 true) (approach-1 event-165486 jack) (approach-2 event-165486 jill)`
- `(touch event-165633 true) (touch-1 event-165633 jill) (touch-2 event-165633 jack)`

The source event (move-inside) is made up of eight micro-events. The event starts with two visible participants, of which one is standing still. The distance between both objects becomes smaller as one participant moves to the other. This continues until they 'touch' each other after which the moving participant disappears out of sight.

- `(visible event-161997 true) (visible-1 event-161997 jill)`
- `(visible event-161791 true) (visible-1 event-161791 house-1)`
- `(move event-161794 false) (move-1 event-161794 house-1)`
- `(distance-decreasing event-162441 true)`
`(distance-decreasing-1 event-162441 jill)`
`(distance-decreasing-2 event-162441 house-1)`
- `(touch event-161801 false) (touch-1 event-161801 jill)`
`(touch-2 event-161801 house-1)`
- `(touch event-162493 true) (touch-1 event-162493 jill)`
`(touch-2 event-162493 house-1)`
- `(borderscreen event-162377 false) (object-1 event-162377 jill)`
- `(visible event-162665 false) (visible-1 event-161997 jill)`

Analogy is commonly defined as a mapping from a source domain to a new target domain. The next step in the algorithm is therefore to check how the existing role maps onto the target event structure. This can be achieved by selecting all the micro-events of the source event which involve the source filler and map it onto the corresponding micro-events of the target. If the micro-events do not exist in the target event, they are ignored. Other information such as time-stamps and hierarchical structure is also ignored. This yields the following mapping:

source event	**==>**	**target event**
(touch-1 event-162493 jill)	==>	(touch-1 event-165633 jill)
(touch-1 event-161801 jill)	==>	(touch-1 event-165633 jill)

A good mapping. There are two requirements for a mapping to be good. The first is that the source filler must always map onto the same object in the target structure. This is indeed the case in the above example: ‘jill’ always maps onto ‘jill’ in the target event. The algorithm thus found an analogy between the source role and a role in the target event. The second requirement for a good mapping is that this corresponding role is the same one as the target role. Again, this is the case in the example: ‘jill’ was indeed the target filler of the target role ‘walk-to-2’. The speaker can thus decide that the existing marker *-ma* can be reused.

In this example, the speaker does not have any other markers yet to check for analogy. If there would be competing analogies, **type frequency** decides which analogy will be chosen (i.e. the semantic role which covers the most participant roles, ranging from one to many). I follow the same definition of type frequency as Bybee & Thompson (2000, p. 77):

> [T]ype frequency determines productivity: type frequency refers to the number of distinct lexical items that can be substituted in a given slot in a construction, whether it is a word-level construction for inflection or syntactic construction specifying the relation among words. The more lexical items that are heard in a certain position in a construction, the less likely it is that the construction will be associated with a particular lexical item and the more likely it is that a general category will be formed over the items that occur in that position. The more items the category must cover, the more general will be its critical features and the more likely it will be to extend to new items. Furthermore, high type frequency ensures that a construction will be used frequently, which will strengthen its representational schema, making it more accessible for further use, possibly with new items.
>
> As type frequency can range from one to a very large number, so there are varying degrees of productivity associated with ranges of type frequency.

In other words, the more general a semantic role, the larger its productivity (i.e. its likelihood to be used in novel expressions).

Adapting the inventory. If no analogy can be found, the agent will invent a new marker as in baseline experiment 2. In this example, however, the agent already knows a suitable marker and it will have to incorporate this new use in its inventory. There are basically two options: either a new verb-specific construction is created for 'walk-to-2' featuring the same case marker as the construction for 'move-inside-1'; or the use of the existing construction is extended. In this experiment, the latter solution is tested.

The changes to the inventory are schematised in Figure 6.10 and can be summarised as follows: the specific meaning in the semantic pole of the construction is removed and replaced by a semantic frame which contains the generalised semantic role. This semantic role shares the same variable as the referent of the other argument unit which was already present in the construction. The two lexical entries which have to be compatible with the new construction (*move-inside* and *walk-to*) are extended with a semantic frame as well. As explained in Chapter 5, this is not a frame in the traditional sense but rather a list of the potential valents of the verb. Since that chapter also gives a full trace of parsing and production, I will not repeat the same operation here.

Learning. The hearer learns the marker by following the same strategy as before. If she didn't know the marker yet, she will create a new verb-specific construction if the context is clear enough. If she already knew the marker, she will get into trouble during parsing because its present use does not correspond to its previous function. The hearer will ignore the problem for the time being and parse the utterance as good as possible. Using the parsed meaning, inferred variable equalities and re-entrance, the agent can then (possibly) retrieve the analogy introduced by the speaker. If the hearer cannot retrieve any analogy, she will nevertheless accept it as imposed by the speaker.

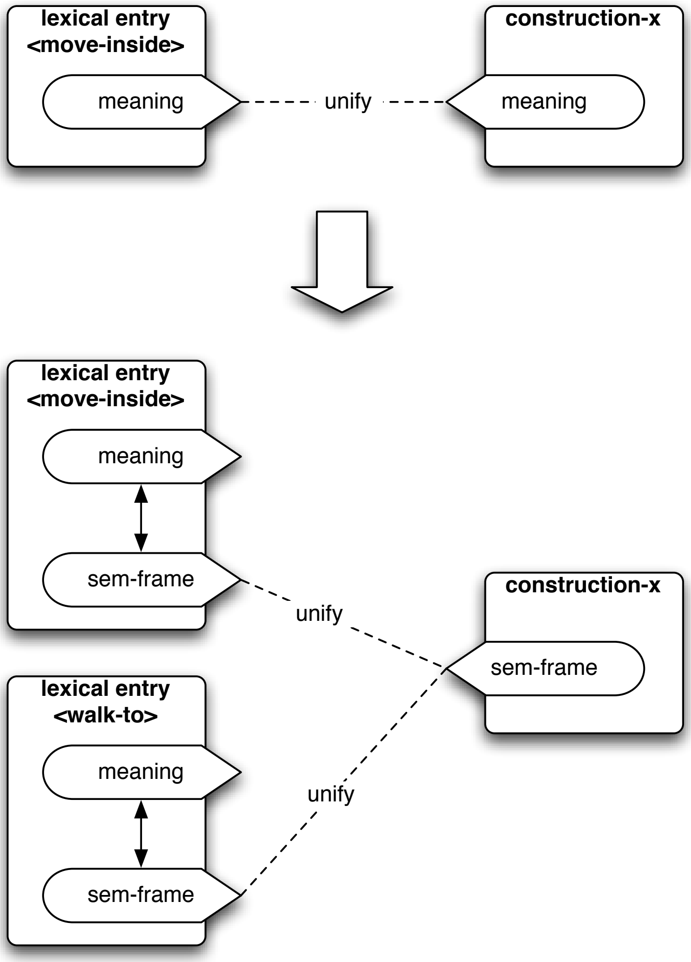


Figure 6.10: This diagram shows how the semantics of lexical entries and constructions integrate with each other. At first, constructions are verb-specific and unify with the meaning of a lexical entry. If the agent however decides to reuse this construction in a new situation, a generalised semantic role is constructed. The relevant lexical entries are extended with a potential semantic frame which unifies with the semantic frame of the construction. The links between the meaning and the semantic frames are taken care of by variable equalities.

The agents in this experiment can thus be characterised as (incremental) instance-based learners (and innovators) (Mitchell, 1997, chapter 8): the agents are 'lazy' learners in the sense that they postpone generalisation until new instances have to be classified as opposed to 'eager' learners that try to make abstractions over the data immediately. Each innovation or novel classification is not based on abstract rules but by examining its relation to previously stored instances. This kind of learning (and innovation) fits usage-based models of language which presuppose *"a bottom-up, maximalist, redundant approach in which patterns (schemas, generalizations) and instantiations are supposed to coexist, and the former are acquired from the latter"* (Daelemans & Van den Bosch, 2005, p. 20, also see section 2.4.3).

Consolidation. The original two-agent simulations did not need to care about alignment strategies or consolidation too much since the two agents always share the same communicative history. So the replicating experiment should pose no problems in both the alignment of case markers and the alignment of the internal structure of semantic roles. The same prediction cannot be made for multi-agent simulations in which alignment strategies are needed for convergence. Three additional set-ups have therefore been implemented: one which uses the same mechanism for updating the confidence scores of linguistic items as in baseline experiment 2 (set-up 3b), one in which a more fine-grained scoring mechanism has been implemented (set-up 3c), and finally one in which (token) frequency decides on the speaker's behaviour (set-up 3d). In this section I will not go into the reasons for experimenting with these different set-ups: they have been inspired by the experimental results and are therefore discussed later on. Instead, I will restrict myself to explaining the two new consolidation strategies. The four different set-ups (and their effects on the results) are summarised in Table 6.3.

Set-up 3c. The more fine-grained scoring mechanism implemented in set-up 3c is based on the idea that linguistic items are not 'good or bad', but that they may be more suitable in some particular contexts and less suitable in others. A single confidence score for every linguistic item cannot go beyond its black-or-white updating scheme and thus cannot handle a more nuanced way of processing. Instead, agents need more clever self-assessment criteria: next to communicative success, they can use **co-occurrences** of linguistic items as a source for aligning their inventories.

Co-occurring items are locally observable to the agents since they form one chain in the reaction network during processing. The general idea is reminiscient of Hebbian learning ('what fires together, wires together'): a link is kept between co-occurring items and a confidence score is kept for this link based on the successful co-occurrence of both items. Suppose that the agent has the case marker *-ma* (see the example earlier in this section) which may cover either the participant role walk-to-2 or the role move-inside-1. The idea is now that the agents keep a link between co-occurring linguistic items, so the agent would now have a link between construction-x one the one hand and the two lexical entries *walk-to* and *move-inside* on the other. Instead of positing a score on the complete construction, each co-occurrence link has its own confidence score:

<Construction-x> ←	0.5 →	<move-inside> (for move-inside-1)
	0.5 →	<walk-to> (for walk-to-2)

Suppose that the speaker also has the marker *-bo* which can be used for marking 'move-1' and 'move-inside-2':

<Construction-y> ←	0.5 →	<move-inside> (for move-inside-1)
	0.5 →	<move> (for move-1)

If the agent then observes the co-occurrence of construction-x and *move-inside* (i.e. the agent analyses an utterance in which *-ma* was used for marking move-inside-1), the score of the link is increased with 0.1 and the score of competing links (here the link between *move-inside* and construction-y) is decreased by 0.1. The other confidence scores based on co-occurrence remain untouched:

<Construction-x> ←	0.6 →	<move-inside> (for move-inside-1)
	0.5 →	<walk-to> (for walk-to-2)
<Construction-y> ←	0.4 →	<move-inside> (for move-inside-1)
	0.5 →	<move> (for move-1)

Note that this score is not the actual co-occurrence frequency, but a confidence score between 0 and 1 which only indirectly reflects co-occurrence and which is updated based on communicative success.

Set-up 3d. The fourth set-up in baseline experiment 3 removes the explicit lateral inhibition consolidation of the previous set-ups and replaces it with a combination of **token frequency and memory decay**. Frequency is implemented as a simple counter which can be updated after each interaction. This set-up has the following features:

- During production, the speaker will use the linguistic items which have the highest frequency score;
- After each *successful* interaction, the hearer will increase the counter of all the constructions that were applied during parsing by one;
- When an agent has engaged in 50 interactions, all the frequency scores are decreased by one (= memory decay).

This kind of (token) frequency favours more general constructions: the larger the type frequency of a certain class or category, the more chances it has to increase its token frequency. The memory decay implemented here is unaffected by population size since it is based on each agent's individual history. It is, however, sensitive to inventory size and frequency: linguistic items can only survive if they occur at least once before the next decay is performed. In the present set-up, each participant role occurs one time out of thirty interactions on average.

6.4.2 Results and discussion of set-up 3a

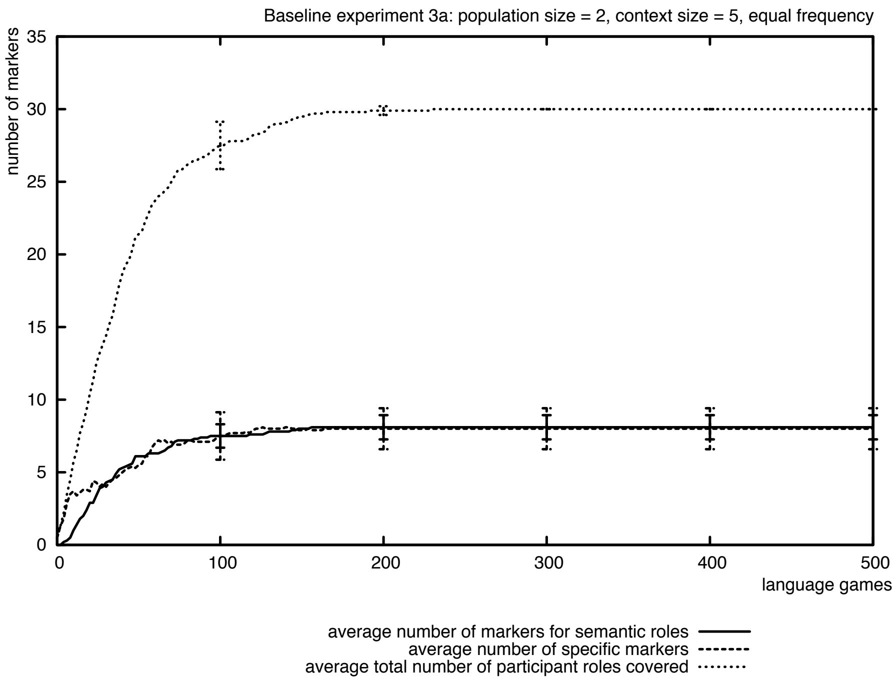


Figure 6.11: In the two-agent simulations, the agents have no problems aligning their inventories since there is no variation in the population. In ten series of 500 language games, the agents came up with an average of 6-8 markers for semantic roles and an average of 6-9 markers for specific participant roles. The semantic roles gather together up to 24 participant roles out of 30.

Results. The replicating experiment featuring a population of two agents confirms the results obtained by Steels (2002b, 2004c). The agents succeed in reusing existing markers and generalising them to semantic roles as is shown in Figure 6.11 . During ten series of 500 language games, the agents constructed on average 6 to 8 markers which could be used for covering at least two participant roles. In total, up to 24 participant roles out of 30 were grouped together in more general roles. In each simulation, also 6 to 9 markers survived which cover specific participant roles.

A closer examination of the semantic roles learns us that they tend to be small generalisations mostly covering two participant roles. Some roles exceptionally gather four or even six participant roles. Here are some example sentences from one of the simulations and their glosses:

(6.20) *jack -cui walk-to jill -ge*
jack sem-role-6 walk-to jill sem-role-26
'Jack walks to Jill.'

(6.21) *touch jill -cui house -shae*
touch jill sem-role-6 house-1 sem-role-29
'Jill touches house-1.'

(6.22) *house -lu move-inside boy -cui*
house-1 sem-role-10 move-inside boy sem-role-6
'The boy moves inside house-1.'

In the same simulation, the following markers and their corresponding participant roles were constructed (ranging from 7 specific markers to 8 more general ones):

- -vuh: cause-move-on-1
- -yaem: cause-move-on-2
- -jibui: cause-move-on-3
- -shuip: give-3
- -vot: take-3
- -me: visible-1
- -naez: move-outside-2
- -zo: fall-1, approach-1
- -tui: fall-2, approach-2
- -shae: touch-2, give-2
- -fe: distance-decreasing-1, grasp-1
- -lu: move-inside-2, distance-decreasing-2
- -we: move-1, give-1, take-1
- -cui: walk-to-1, object-1, grasp-2, hide-2
- -ge: touch-1, move-inside-1, move-outside-1, hide-1, walk-to-2, take-2

Discussion. The results show that the construction of generalised semantic roles allows the agents to reduce the number of markers by 65–70%. However, the most important observation is that by endowing the agents with the capacity of analogical reasoning, they are capable of generalisation beyond previous linguistic experience as is shown in the increasing productivity of some markers.

In the results there is still a fairly large residue of verb-specific markers which is partly due to the analogy algorithm and partly due to the fact that only two agents were involved in the simulation. First, the analogy algorithm is very strict in the sense that two roles are either analogous or not: there is no in-between value that allows for some flexibility. Second, since there are no variations in the population, the construction of semantic roles is entirely dependent on the linguistic history of both agents: once an analogy is constructed and successfully applied in communication, the agents will not try to come up with better or more general analogies later on. In other words, the solutions that the two agents come up with may not be optimal given their search space so they end up in a local maximum. A larger population may give this search an additional boost.

6.4.3 Results and discussion of set-up 3b

Results. In set-up 3b the population size is increased to 10 agents so there will be more variation among the agents. This is indeed confirmed in Figure 6.12 which shows that there are a total of 140 variations floating in the population for marking 30 participant roles. This is an average of 4,7 possible ways for marking each participant role. We see that this number of possibilities does not drop to 30 which is a first indication that the agents do not converge on a shared set of grammatical markers. With respect to the nature of the markers, the results indicate that there are about 20 markers which can be used for covering at least two participant roles whereas there are about 50 specific participant role markers as well. The average inventory size of the agents is thus far from optimal.

Figure 6.13 confirms that the agents do not reach convergence: the meaning-form coherence indicates the degree to which the agents prefer the same case marker for a particular participant role. As the graph shows, coherence only reaches 40% which means that the agents use a different marker for the same participant role in more than half of the language games. Yet, as the graph also shows, the agents are capable of reaching 100% communicative success and reducing the cognitive effort to zero.

Discussion. The results of baseline experiment 3b seem to be contradictory at first sight: even though the agents do not agree an a shared preferred set of markers, they nevertheless reach communicative success. This is possible because success in communication does not require meaning-form coherence: if the agents learn all the variations floating around in the population, they can still parse all utterances correctly. This is indeed what happens in this experiment.

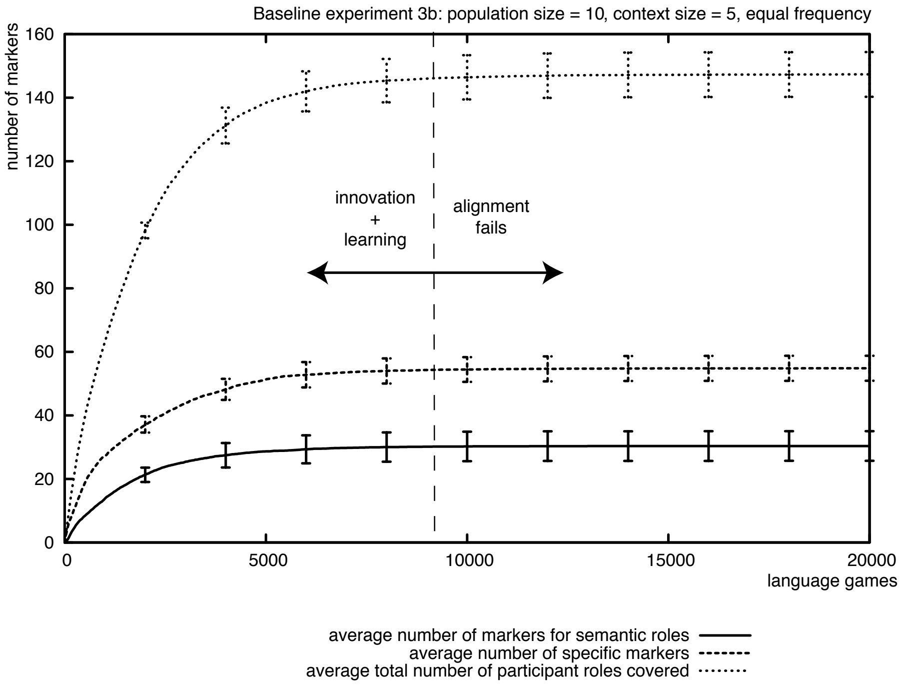


Figure 6.12: When scaling the experiments up to multi-agent simulations, the traditional alignment strategies used in prior experiments on lexicon formation are not sufficient for the population to reach convergence. For thirty participant roles, the agents have to remember about 140 variations in order to communicate successfully, which is an average of 4,7 possible ways for marking each participant role.

The lack of convergence on a preferred set of markers clearly indicates that the proposed alignment strategy is insufficient. The reason is that the alignment strategy in which each item has its own single confidence score is best suited for simulations in which there is always a *one-to-one* mapping between form and meaning (as was the case in baseline experiment 2). However, when markers get generalised to cover more than one participant role, they become polysemous *one-to-more* mappings.

I will go through an example to explain why the single confidence score cannot be sufficient for polysemous form-meaning mappings. Suppose that an agent knows three markers: *-ma*, *-bo* and *-li* and that the first two are generalised to cover two roles each whereas *-li* is still a verb-specific marker. For convenience's sake, I will not include all the linguistic items involved but treat the markers as if they were lexical items:

$$\begin{array}{ll} \text{<move-inside-1>} & \leftarrow \\ \text{<walk-to-2>} & \leftarrow \end{array} (0.5) \rightarrow \text{-ma} \tag{6.23}$$

$$\begin{array}{ll} \text{<move-inside-1>} & \leftarrow \\ \text{<move-1>} & \leftarrow \end{array} (0.5) \rightarrow \text{-bo} \tag{6.24}$$

$$\text{<move-1>} \quad \leftarrow (0.5) \rightarrow \text{-li} \tag{6.25}$$

Suppose now that the agent observes the utterance *boy -ma move-inside* in which the marker *-ma* was successfully used for marking ‘move-inside-1’. The score for *-ma* is thus increased and the score for its competitor *-bo* is decreased. The consequences for *-bo* are far-reaching, because it is now not only less successful than *-ma* for covering ‘move-inside-1’, but also than *-li* for marking ‘move-1’:

$$\begin{array}{ll} \text{<move-inside-1>} & \leftarrow \\ \text{<walk-to-2>} & \leftarrow \end{array} (0.6) \rightarrow \text{-ma} \tag{6.26}$$

$$\begin{array}{ll} \text{<move-inside-1>} & \leftarrow \\ \text{<move-1>} & \leftarrow \end{array} (0.4) \rightarrow \text{-bo} \tag{6.27}$$

$$\text{<move-1>} \quad \leftarrow (0.5) \rightarrow \text{-li} \tag{6.28}$$

In another game, the same damage can be done for the marker *-ma* so *-bo* can recover from its score decrease. Generalisation thus tends to be harmful for the markers if only one score is used: the more general a role marker gets, the more competitors it has and thus the more chances that its score will be decreased. Specific markers can escape punishment through lateral inhibition much more easily. On the other hand, if it has competing markers which are generalised, they can get punished too even if a different participant role was involved. Suppose that *-bo* was observed for marking ‘move-inside-1’ this time, then not only *-ma* is seen as a competitor, but also *-li* because it overlaps with *-bo* for marking ‘move-1’.

There is thus a constant push-and-pull effect in which markers may get cornered by others but then all of a sudden get more successful again. This is the reason why the agents can never converge on a preferred set: the single confidence score does not allow markers to be successful in one particular context but unsuccessful in another.

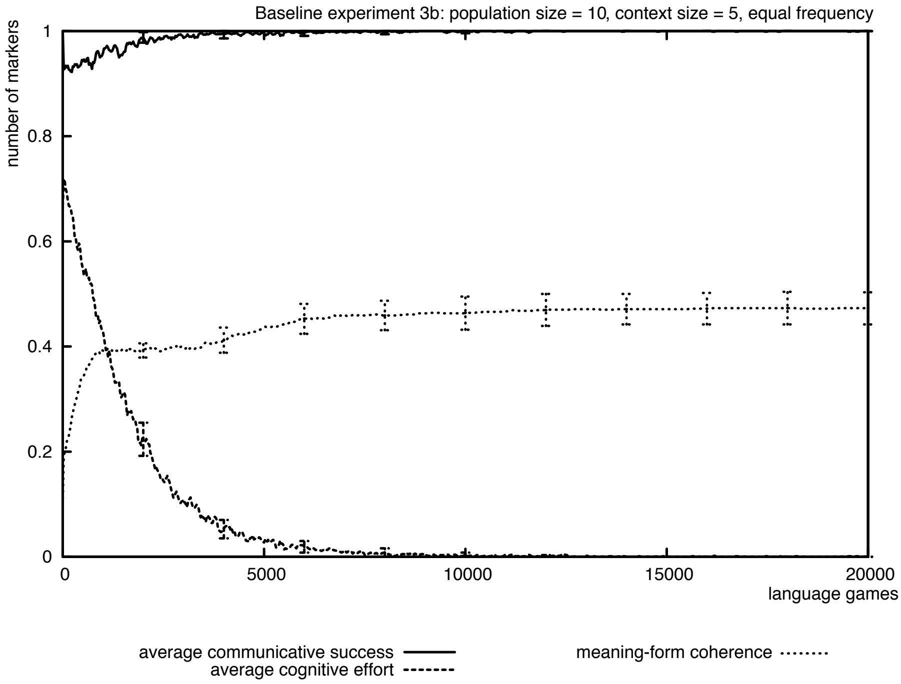


Figure 6.13: This graph shows that the agents reach 100% communicative success and reduce the cognitive effort needed for interpretation. However, the meaning-form coherence only reaches about 40% which indicates that the agents did not converge on a shared set of markers but keep using divergent preferences in more than half of the language games.

6.4.4 Results and discussion of set-up 3c

Results. The results of baseline experiment 3c indicate that the alignment strategy of reinforcement and lateral inhibition can lead to convergence if it is applied in a more fine-grained way. As explained before, the agents this time not only use communicative success as a guidance but also co-occurrence links: instead of positing one score on the linguistic item as a whole, they now keep a link between co-occurring items and assign a confidence score to that link. In case of success, the score of the link is increased and only scores of competing *links* are decreased. In this model, a marker disappears from the linguistic inventory once it has no links to other linguistic items anymore with a confidence score higher than zero.

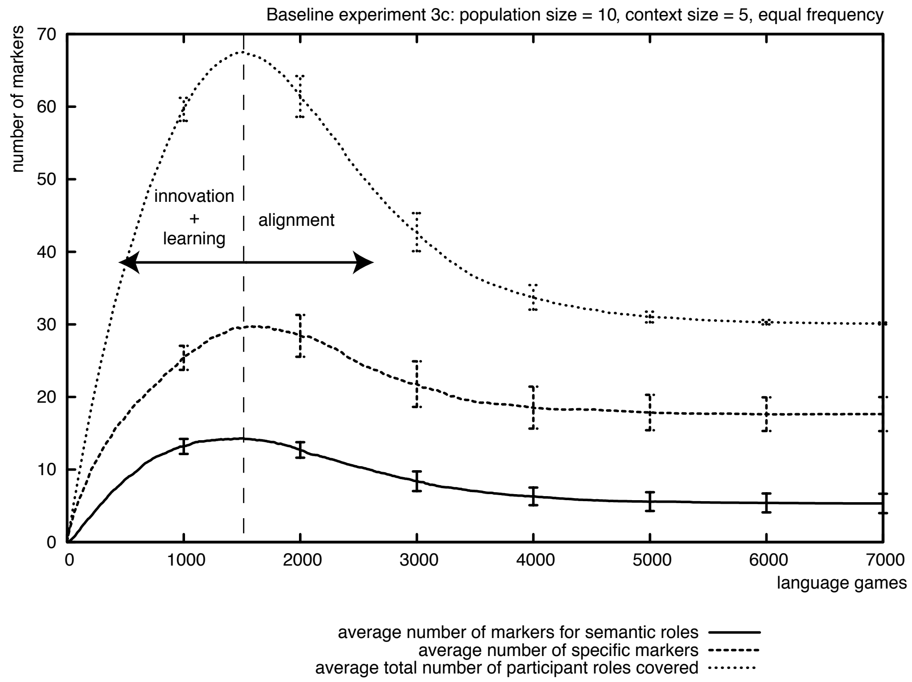


Figure 6.14: The more fine-grained alignment strategy allows the agents to converge on one possible marking for each of the thirty participant roles. There are 5 semantic roles on average which each cover about two participant roles. The remaining 20 roles are covered by specific markers.

Figure 6.14 shows that the number of variations peaks at 70 possible markings for 30 participant roles. This means that the agents only have to deal with an average of 2,3 competing markers for each participant role. Innovation and adoption of markers stops at about 1.500 language games after which the agents rapidly converge on a shared set of markers. The graph also shows that the agents converge on a set of 5 generalised semantic role markers and about 20 verb-specific markers. This means that the verb-specific markers managed to win the competition more often than generalised markers and that semantic roles on average only cover two participant roles.

Figure 6.15 shows that the fine-grained alignment strategy allows the agents to converge on a shared set of preferences: meaning-form coherence reaches 100% after 5.000 language games which corresponds to the moment where the agents have pruned all the variations down to 30 in Figure 6.14. The agents also reach communicative success and manage to reduce the cognitive effort needed during parsing.

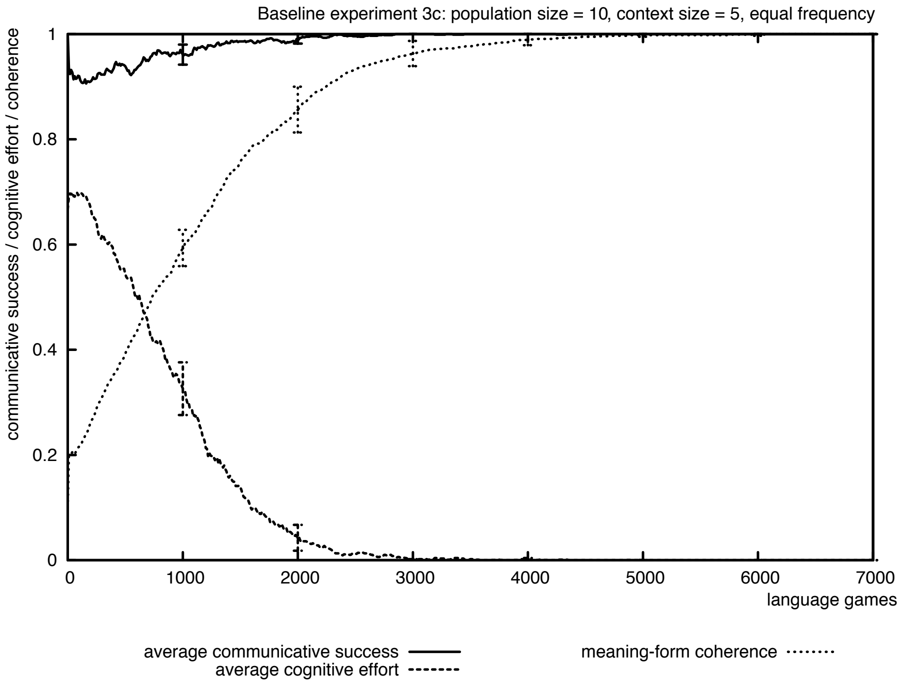


Figure 6.15: Using the more fine-grained alignment strategy, the agents not only succeed in reaching communicative success and reducing cognitive effort, they also converge on a shared set of meaning-form conventions.

Discussion. The results show that the fine-grained scoring mechanism suffices to solve the problem of convergence among the agents. However, the gain in inventory optimisation is minimal: the agents end up with an average of 25 markers for 30 participant roles. Also the benefits of generalisation are on the low side with an average of two participant roles covered by a semantic role.

By solving the problem of the single confidence scores, the fine-grained scoring mechanism created a new one: since only competing links are taken into account during consolidation, the influence of the frequency of the entire category is neglected. This means that a verb-specific marker has the same chances of surviving the competition as generalised semantic role markers do, even though the latter ones are as a whole more frequent and productive. If a semantic role loses the competition from a specific marker, its type frequency is reduced and hence its productivity.

To overcome this problem, the agents need another alignment strategy which both recognises the impact of generalised roles and is capable of dealing with the context-sensitive nature of polysemous markers. Experiment 3d implements such a strategy.

6.4.5 Results and discussion of set-up 3d

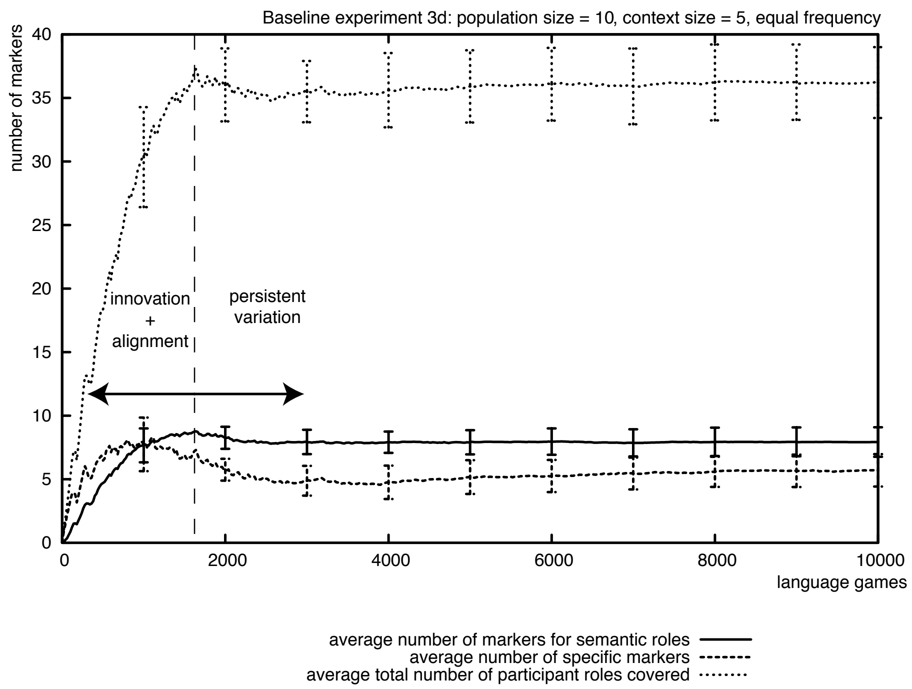


Figure 6.16: When adapting their linguistic behaviours to frequency, the agents tend to use more generalised semantic role markers rather than specific ones. Here, about seven semantic roles cover 25 of the 30 participant roles. The top line indicates that some amount of variation persists over time.

Results. The final set-up in baseline experiment 3 does not use confidence scores or lateral inhibition. Instead, agents rely on token frequency of successful interactions for producing utterances. Figure 6.16 shows that the agents spend roughly the same time as in set-up 3c innovating and learning new markers. The average amount of variation reaches a total of 35–40 possibilities for 30 markers, which is an average of less than two variations for each participant role. The innovation rate is in fact as high as in the previous set-up, but many innovations are excluded very early on by memory decay. Innovations that do survive the memory decay during the first 2.000 interactions are quite frequent so they persist in memory for a very long time afterwards. Getting rid

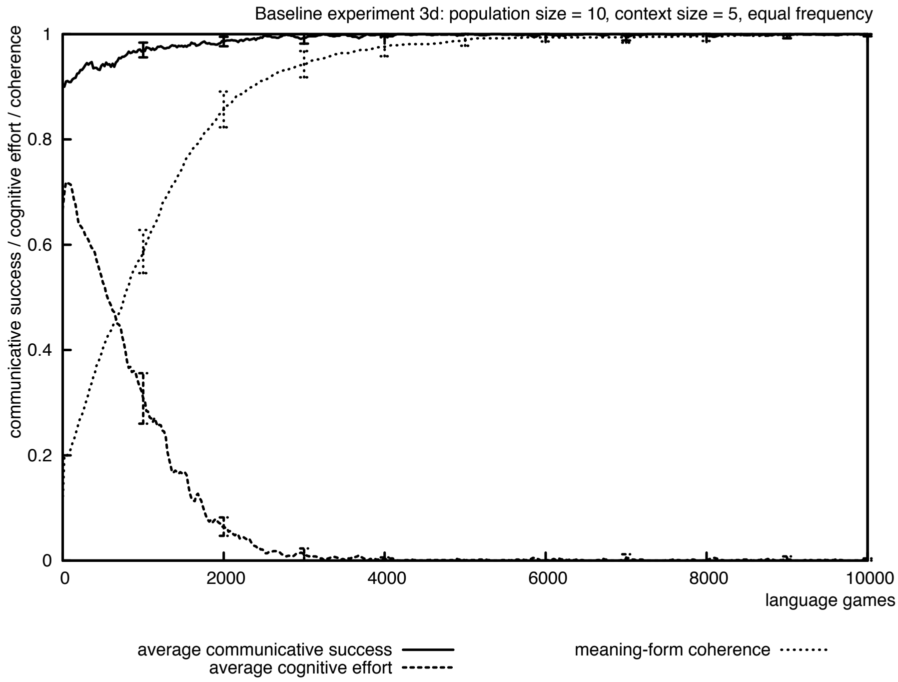


Figure 6.17: This graph shows that the agents reach communicative success, reduce the cognitive effort needed for interpretation and converge on a shared set of case markers.

of them is therefore very slow and may take thousands of additional language games before they are ‘forgotten’ or in some cases they persist over time. A closer look at the markers themselves learns us that there are on average eight semantic roles and six markers for specific participant roles. This means that the semantic role markers can cover up to 24 participant roles.

Figure 6.17 indicates that even though the agents do not reduce their grammars to a single variation for all 30 participants, they nevertheless converge on a shared set of preferred markings: coherence rises to 100% in 6.000 language games. Communicative success reaches 100% and the agents rapidly succeed in reducing the cognitive effort needed for interpretation.

Discussion. In order to interpret the results of set-up 3d correctly, a closer examination of the artificial languages of the agents is needed. In one of the simulations, a population of ten agents all preferred the following 14 markers and their corresponding participant roles:

- *-zoti*: cause-move-on-1
- *-ruko*: cause-move-on-2
- *-jaexi*: grasp-2
- *-mad*: approach-1
- *-zima*: give-1
- *-wobae*: give-2
- *-ha*: take-3
- *-qui*: cause-move-on-3, visible-1
- *-fechui*: touch-2, take-1
- *-kuwae*: touch-1, take-2
- *-yuis*: fall-2, approach-2
- *-pae*: give-3, walk-to-2, move-outside-1
- *-ru*: walk-to-1, distance-decreasing-2, move-1, move-inside-2, hide-2, move-outside-2
- *-gahu*: object-1, move-inside-1, fall-1, distance-decreasing-1, hide-1, grasp-1

The above markers suggest that there were seven specific markers left and seven semantic role markers. However, the marker - *jaexi* can in fact be counted as a semantic role because it can also cover the participant roles 'hide-2' and 'distance-decreasing-2'. In both cases, however, the marker is in competition with *-ru* which has both a higher type and token frequency. Similarly, *-pae* can also cover 'hide-1' but this participant role is dominated by the frequent role marker *-gahu*. This explains why in this simulation there remain 33 possibilities for 30 participant roles instead of only 30: the markers *-jaexi* and *-pae* have found their own 'semantic niche' in which they occur frequently enough to avoid memory decay. Synchronic variation like this is in fact more realistic than the competition dynamics in the previous set-ups since it causes a pool of variation which may trigger future changes in the language: both markers may disappear after a while or they may extend their usage and become stronger rivals for the now more successful markers *-ru* and *-gahu*.

When comparing the results to the two-agent simulations, set-up 3d improves in terms of generalisation: more participant roles are covered by the same marker. The improvement is however not that big so the simulations do not demonstrate that the collective solution found by larger populations can avoid the local maxima that two agents encountered in their communicative interactions. In order to fully test this hypothesis, experiments are needed involving a larger and more controlled search space.

The alignment strategy however does succeed in favouring the more general roles through function and frequency: markers which have a higher type frequency and therefore a wider usage tend to have a higher token frequency as well. This creates the same rich-get-richer dynamics of the strategy involving one score and lateral inhibition in baseline experiment 2: the more frequent a marker is, the more likely it will win the competition in the future and the more likely it will increase its type frequency as well. At the same time, the alignment strategy allows for the same context-sensitivity as the fine-grained scoring mechanism because it does not feature explicit lateral inhibition so no categories are unrightfully harmed by it. This allows more lexical markers to still survive in their (sometimes verb-specific) semantic 'niche' if they are frequent enough to survive memory decay.

6.4.6 Conclusions and future work

In this section I discussed the various set-ups of baseline experiment 3 and reported on its results. Common to all simulations was the additional cognitive ability of analogical reasoning over event structures. This cognitive mechanism allowed the agents to reuse (and generalise) existing markers in new situations and contexts. By exploiting analogy, the agents are thus capable of generalising their grammars beyond the input of previous experiences. Generalisation is thereby not a goal in itself, but rather a side-effect of the need for optimising communication in an inferential coding system.

Four set-ups were implemented and compared to each other. The first set-up successfully replicated the original case experiment and set the baseline for the other three multi-agent simulations. Set-up 3b indicated that a single confidence score is not a sufficient alignment strategy for converging on a shared set of preferred markings: this strategy is optimised for one-to-one mappings but cannot deal with the context-sensitiveness of polysemous one-to-many mappings. An alternative was therefore implemented in set-up 3c in which the agents also exploited the co-occurrences of linguistic items: this time, a specific link was kept between all co-occurring items with a confidence score for each link. The strategy proved itself sufficient for reaching coherence in the population but at the cost of generalisation. Finally, an alignment strategy was proposed based on token frequency and memory decay. This strategy led the agents to convergence on a set of preferred markings and improved slightly over the results of the two-agent simulations. The various set-ups are summarised and compared to each other in Table 6.3.

In the simulations, analogy is the source of generalisation and increased productivity of existing markers. This suggests that (at least for innovation and learning), analogy can be used as a unified account for both the more 'regular' forms in the language and the more 'irregular' forms as opposed to rule-based accounts which posit abstract rules and a list of exceptions. In order to exploit the power of analogy, however, the agents need the right kind of alignment strategy that favours the more general categories.

Exp.	**Agents**	**consolidation**	**Effect**
3a	2	store innovations	–
3b	10	store innovations + confidence score on all items + lateral inhibition	alignment fails
3c	10	store innovations + confidence score on co-occurrence + lateral inhibition	alignment succeeds → arbitrary winners
3d	10	store innovations + frequency of constructions + memory decay	alignment succeeds → general roles favoured

Table 6.3: This table compares the four alignment strategies implemented in baseline experiment 3. In set-up 3a, no additional alignment strategies were needed since there were only two agents and hence no variation was observed in the population. Set-up 3b showed that direct competition did not yield successful alignment because a confidence score on each linguistic item cannot deal with polysemous usage of the items. Set-up 3c solved the problem of alignment through a confidence score on each co-occurrence link. This strategy however led to equal opportunities for each marker in the population so unproductive markers survived as easily as general ones. The final strategy involved the frequency of construction tokens which favoured more general markers because they have wider application and are thus more frequent.

Analogy all the way. Evidence from natural languages suggests that analogy is also responsible for the first innovation by recruiting an existing lexical entry for a more grammatical use instead of inventing a new marker (see section 4.2.3). Additional experiments are thus needed in which the innovation strategies of baseline experiments 2 and 3 are combined into one. The recruitment of existing and well-entrenched lexical entries would naturally follow from the same assumptions that language is an inferential coding system and that speakers and hearers will exploit whatever resources that are available for solving communicative problems: using a conventionalised linguistic unit has the major advantage that it offers the hearer a strong grounding point for inferring the meaning or function of the innovation. In the present experiments, the new markers do not contain any clues about what their source events were, so this may differ strongly from agent to agent which explains why the hearer can't retrieve the analogy in all cases. In the abstractions and scaffolds of the present set-up, however, this poses no problems.

Implementing a more realistic model of stage 2 in the development of case markers, however, is not a trivial matter and requires more study on how this happens in natural languages. From the evidence gathered so far (see section 4.2.3), it seems that the present algorithm for analogy cannot handle this. The first reason is that the recruited lexical item in serial verb language constructions, which are typical sources of case markers, seems to 'fit' naturally in the utterance by for example presupposing the same subject as was demonstrated in example 4.40 which I repeat here:

(6.29) *thân cà bin maa krungthêep*
he will fly come Bangkok
'He will fly to Bangkok.'
(Blake, 1994, p. 163)

The present analogy algorithm already expects a marker and would not know how to deal with the other participant roles of the recruited verb. The task can involve even three participants in cases where for example *give* evolves into a dative or recipient marker. The data thus show that next to a more general-purpose algorithm for analogy, we also need to find solutions for coordination and ellipsis so that the recruited lexical entry can naturally blend in the utterance. Before we can do this, however, we first need to investigate how the syntactic categories can be formed that have to be coordinated.

A second problem has to do with morphology and phonological reduction. In the attested examples, the second verb in a serial verb language construction is implicitly marked for its more grammatical function because it typically occurs in a non-finite or a non-conjugated form. In the experiments, there is no morphology or syntax that could distinguish two verbs from each other so the hearer would have a very hard time at figuring out which verb was meant as the 'main verb' and which one was meant as the 'marker'. Moreover, the hearer would have no reasons to assume that one of the verbs has been recruited for a new use in the first place. The problem with phonological reduction is that there is no phonological component in the experiments so recruited lexical items cannot evolve towards a new form which distinguishes them more clearly from their original uses.

Next to work on syntax, coordination and morphology, a dynamic representation of categories and word meanings is needed. First steps have already been taken by Wellens (2008) who investigates how word meanings can become more flexible and therefore change over time. This work however only deals with words for objects so more effort is needed to integrate it with the architecture of the experiments in this thesis. Another particular issue with the model of Wellens is that it does not allow true polysemy: the agents continuously shape the meaning of a lexical entry but they cannot use the same word in multiple ways. The meanings are therefore still one-to-one mappings between form and meaning, but what the exact content of the 'one' meaning is may change over time. Grammaticalization of case markers, however, requires one-to-many mappings or even many-to-many mappings, so the agents have to be capable of distinguishing between different uses of the same form. I believe that coordination and pattern formation could be a key in solving this issue, as I will explain in more detail in section 8.4.3.

6.5 Summary: towards argument-structure constructions

This chapter presented three baseline experiments which build on original work by Steels (2002b, 2004c). After explaining the basic set-up of the experiments, the original two-agent simulations were successfully replicated. Each baseline experiment was then scaled up to a multi-agent simulation involving a population of ten agents. For each experiment, the necessary diagnostics, repairs and alignment strategies were defined which were shown to be crucial for the transition from one stage to the next. The experiments reported in this chapter are the first successful multi-agent simulations reported in the field that feature polysemous categories.

Baseline experiment 3 illustrated the kind of usage-based model that I introduced in section 2.4.3. In this model, careful abstraction is allowed so that the generalisation and performance accuracy of the agents could not get harmed. This redundant, bottom-up approach of careful abstraction is a robust way of making generalisations even though the conventions of the population are not fixed yet. I demonstrated that analogy is not only useful for extending the use of existing linguistic items, but that it can also be the cause of increased productivity. This view is most closely related to the usage-based model as proposed by Ronald Langacker:

> I suggest that extension tends to be accompanied by schematization, that the "outward" growth of a network by extensions from a prototype tends to induce its "upward" growth via the extraction of higher-level schemas. (Langacker, 2000, p. 12)

So far, the experiments feature case markers which developed (mostly) independently of each other. As argued in Chapter 4, however, case markers can only be fully understood if they are examined in relation to each other. In other words, the next step is to allow the agents to construct larger patterns for marking event structure. This will be the main topic of the following chapter which investigates how larger argument structure constructions can be formed.

Chapter 7

Multi-level selection and language systematicity

The baseline experiments of the previous chapter looked at how analogy could be exploited for the generalisation of case markers for covering semantic roles. The experiments focused on the development of these semantic roles in isolation of each other in order to identify the diagnostics, repairs and alignment strategies that make the formation of such roles possible. However, as I discussed in Chapter 4, the behaviour and functionality of case markers can only be fully understood when they are studied in relation to the other elements in their linguistic context. In other words: case markers have to be investigated in relation to the patterns in which they occur. This chapter therefore presents experiments in which case markers can be combined in larger patterns.

The next section first gives a brief overview of pattern formation in language and operationalises one strategy of pattern formation in the form of diagnostics, repairs and alignment strategies. Section 7.2 implements this operationalisation and shows that the 'systematicity' of the artificial languages gets lost once smaller linguistic units are starting to combine into larger patterns. In this section I will also briefly discuss other experiments in the field in which the problem of systematicity occurs but is either overlooked or misinterpreted by the experimenter. The next section then presents the results of another experiment that uses the more complex alignment strategy of multi-level selection to overcome this problem. Three variations of multi-level selection are implemented and compared to each other in terms of systematicity and coherence. The insights of these experiments are ported to experiments involving analogy and the formation of semantic roles in section 7.4. Section 7.5 finally offers a first step towards simulations involving the formation of syntactic cases (corresponding to stage 3 in section 4.2.4). Even though stage 3 is not fully accomplished yet, this section offers a clear idea of the work that needs to be undertaken in order to form syntactic cases.

7.1 pattern formation

One crucial aspect of grammaticalization (see Martin Haspelmath's definition in section 6.1.4) is the drift towards tighter structures and a lesser degree of freedom. For example, lexical items develop into more grammatical items and become part of (larger) constructions. Within these constructions or patterns, the freedom of the individual parts is restricted and depends on the pattern as a whole. This would explain why for example an allative case marker only makes sense in a motion-pattern as discussed in Chapter 4. However, linguistic items that become part of a larger construction may still have a life on their own in their original sense, a phenomenon traditionally known as 'layering' (Hopper & Traugott, 1993, p. 124–126). For example the preposition *like* can still be used for indicating similarity while at the same time it can be used as a marker for introducing reported speech:

(7.1) She looks nothing *like* her father.

(7.2) And he was *like* "Oh that is so not true!"

In the following subsection I will briefly touch upon some phenomena of grammaticalization involving pattern formation and offer an analysis which is somewhat different from the traditional linguistic approach. I will support my analysis through other examples of patterns and idioms in language. In the next subsection, I will then offer an operationalisation of my analysis in terms of diagnostics and repair strategies for the artificial agents that will be used in the experiments in this chapter.

7.1.1 pattern formation in language

A very good example of the development of a lexical item into a part of a grammatical structure can be found in French negation. Traditionally, the development of negation particles (also known as 'Jespersen's cycle') is defined in terms of a cycle of reanalysis – analogy (generalisation) – reanalysis (Hopper & Traugott, 1993, p. 65–66):

1. Negation in French originally only involved *ne* before the verb:

   (7.3) *Il* *ne* *va.*
   he NEG go.3SG.PRES
   'He doesn't go.'

2. In the context of motion verbs, *ne* could optionally be reinforced by the noun *pas* 'step':

   (7.4) *Il* *ne* *va* *(pas).*
   he NEG go.3SG.PRES (step)
   'He doesn't go (a step).'

3. The word *pas* is reanalysed as a negator particle in the construction [*ne* Vmotion *pas*];

4. The particle *pas* is extended analogically to non-motion verbs as well:

    (7.5) *Il* *ne* *sait* *pas.*
    he NEG know.3SG.PRES NEG
    'He doesn't know.'

5. The particle *pas* is then reanalysed as an obligatory part of the construction [*ne* V *pas*];

6. In spoken French, *ne* is reanalysed to become optional and is eventually lost:

    (7.6) *Il* *sait* *pas.*
    he know.3SG.PRES NEG
    'He doesn't know.'

Reanalysis versus pattern formation. Reanalysis is essentially a hearer-based analysis of this developmental cycle in which the hearer interprets the underlying structure of an utterance in another way than was intended by the speaker. Reanalysis is traditionally understood as *"change in the structure of an expression or class of expressions that does not involve any immediate or intrinsic modification of its surface manifestation"* (Langacker, 1977, p. 58). Even though reanalysis is a plausible mechanism for step 3, its main problem is that it is invisible from the outside. Hopper & Traugott (1993) write that for *"the French negator* pas*, we would not know that reanalysis had taken place at stage [3] without the evidence of the working generalization at stage [4]"* (p. 66). As Haspelmath (1998) points out, however, this means that reanalysis cannot explain how the new use of *pas* got propagated and accepted in the speech community unless all speakers are assumed to make the same reanalysis at roughly the same time which is very implausible. As I will explain more thoroughly in section 8.4.3, reanalysis needs to be accompanied by other mechanisms in order to account for the empirical data.

I propose a different and simpler mechanism for step 3 which is in line with the general approach of usage-based models of language: pattern formation. If a certain group of words occur frequently enough together, they are stored as a new unit in the linguistic inventory. This means that the language user now knows two **competing** constructions in the case of motion verbs: [*ne* V] and [*ne* Vmotion *pas*]. This approach of pattern formation may seem redundant from the point of view of inventory size, but it may optimise linguistic processing because a pattern is a 'pre-compiled' chunk that is readily available for use, whereas otherwise the language user needs to compose the structure over and over again. Since pattern formation is a relatively 'simple' operation for optimising processing, we can assume within a usage-based model that most language users will do this spontaneously for all recurrent patterns in the language as opposed to a collective operation of reanalysis. Once a pattern is stored in memory, it can start a life on its own and diverge from its original usage. Steps 1–5 in the negation cycle can thus be reinterpreted as follows in a more speaker-based analysis:

1. Negation in French originally only involved *ne* before the verb:

   (7.7) *Il* *ne* *va.*
   he NEG go.3SG.PRES
   'He doesn't go.'

2. The speakers of French start to reinforce the negation particle *ne* in some situations to put more emphasis on the negation or to solve communicative problems. In the context of motion verbs, the reinforcement is achieved through the noun *pas* 'step', whereas in other contexts such as verbs of visual perception, negation is reinforced through *point* 'point':

   (7.8) *Il* *ne* *va* *(pas).*
   he NEG go.3SG.PRES (step)
   'He doesn't go (a step).'

   (7.9) *Il* *ne* *voit* *(point).*
   he NEG see.3SG.PRES (point)
   'He doesn't see (a point).'

3. The frequent use of these reinforcement nouns leads to the creation of readily available patterns which co-exist (and compete with) the standard negation construction;

4. The new patterns are extended analogically to non-motion verbs as well and start to compete with each other and with the old negation construction for becoming the new default negation;

5. The construction [*ne* V *pas*] wins the competition and becomes the new default construction for negation. Other competitors using different particles either disappear or take up their own semantic niche (*ne ... point* 'nothing' (old-fashioned), *ne ... plus* 'no more', *ne ... rien* 'nothing', *ne ... jamais* 'never', *ne ... guère* 'almost nothing', etc.). The old negation construction gets lost except for some archaic uses in writing.

Idioms. Evidence for pattern formation as opposed to reanalysis can be found in idioms. Idiomatic expressions have always been problematic for traditional linguistic theories which take a modular approach to language and assume a sharp distinction between conventional-lexical items and systematic-syntactic rules. Faced with such problematic issues, usage-based models and particularly construction grammars *"grew out of a concern to find a place for idiomatic expressions in the speaker's knowledge of a grammar of their language"* (Croft & Cruse, 2004, p. 225). Idioms range from highly idiomatic expressions to more schematic constructions (Croft & Cruse, 2004, chapter 9):

(7.10) by and large; no can do; be that as it may; make believe; so far so good

(7.11) kick the bucket; pull a fast one; spill the beans

(7.12) to answer the door; wide awake; bright red; to blow one's nose

(7.13) the bigger the better; the louder you shout, the sooner they will serve you

No theory of grammaticalization that I am aware of explains idioms such as *so far so good* or *by and large* in terms of reanalysis of the words that make up the idiom. Similarly, compound nouns are given their own lexical entry rather than introducing a notion of 'synchronic layering' (Hopper & Traugott, 1993, p. 124–126) over the original words caused by reanalysis. Also pattern formation on other levels of language (e.g. reoccurring syllables, morphemes, etc.) are never treated as synchronic layers on top of one entry in the linguistic inventory. Reanalysis is therefore used in an ad-hoc way, or as Haspelmath (1998) writes, *"as one pleases"* (p. 341).

By taking pattern formation seriously, meaning that many redundant copies exist in memory, a simpler alternative exists for the ad-hoc mechanism of reanalysis. Just as there is no reason for differentiating 'core case markers' from 'peripheral semantic case markers' (see section 4.2.5), the language user makes no difference between fully idiomatic expressions such as *by and large* and more grammatical constructions such as [*ne ... pas*]. The only difference between them is that the more schematic constructions were extended and generalised to new uses whereas the more idiomatic expressions remained unchanged depending on communicative needs in language use and frequency effects. This usage-based approach naturally leads to the continuum of linguistic items as observed in natural languages.

One problem with the alternative hypothesis is that it is invisible from the outside just like reanalysis is. This is where computational models can prove their worth: they can **demonstrate** the consequences of each alternative hypothesis and show what kind of cognitive apparatus is needed for both. Additional evidence can then be gathered from other disciplines such as psycholinguistics to determine which cognitive architecture is most plausible. So even though computational modeling cannot predict actual language change, they can demonstrate the effects of proposed mechanisms and help to fill in the blanks when there is a lack of empirical data as is the case in grammaticalization theory. I will return to this matter in section 8.4.3.

7.1.2 Operationalising pattern formation

The above idea of pattern formation needs to be implemented in terms of diagnostics and repair strategies that make use of information that is locally available to the agents. Consider the reaction network (i.e. search tree; see section 5.3) of Figure 7.1 in which an agent used two constructions which subsequently licensed node-2 and node-3 in the network and which licenses the utterance *jack -bo push block -ka*:

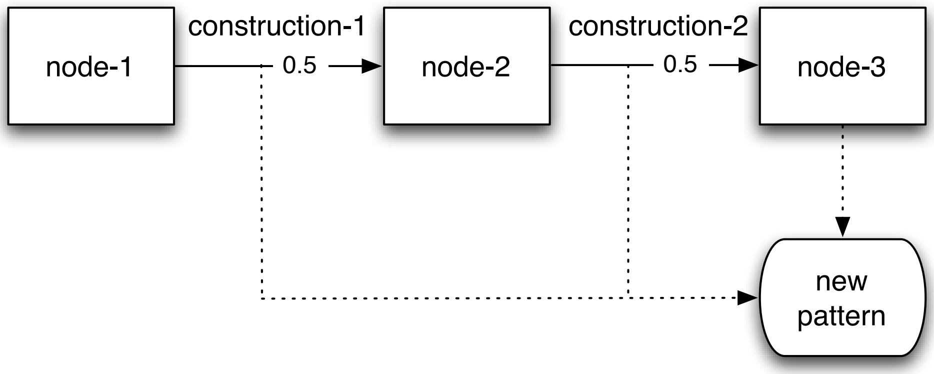


Figure 7.1: An agent's reaction network is the source for pattern formation. If the agents have to apply two constructions to license an utterance (production) or a meaning (parsing), they will create a pattern based on the applied constructions. This pattern has the same functionality as the constructions but only requires one step.

Suppose that the agent is in production mode. In this case node-1 is the coupled feature structure which was licensed after unifying and merging the lexical entries for *jack, push* and *block*. Next, the speaker has to unify and merge two constructions for marking the two participants of the push-event which licenses node-3. In a next step, which is not shown in the figure, the agent will unify and merge the morphological rules. As indicated in the figure, this reaction network forms the basis for a new pattern (which will be construction-3). In principle this pattern should combine the entire reaction network including the lexical entries, but for convenience's sake the agents will only make a pattern which combines the functionality of constructions 1 and 2:

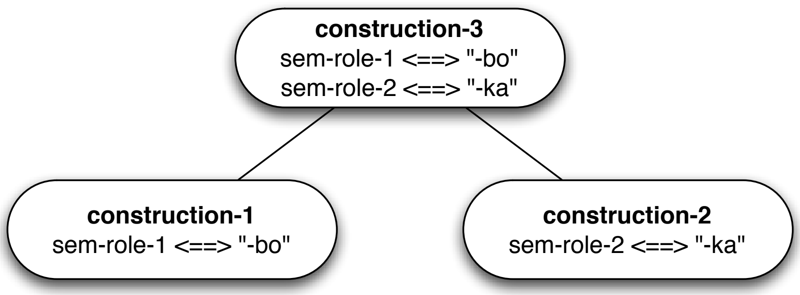


Figure 7.2: The two constructions that were used during processing are combined into a new construction. The agents keep a link between the new construction and the constructions that were used for creating it.

The new construction is stored in the linguistic inventory with information about its origins: the agents keep a link between the new pattern and the constructions that were used for creating it. If the speaker has to produce the same meaning again, the new construction now forms an alternative path in the reaction network. The speaker will prefer this new path because it is faster in processing (one step can be skipped) and the links between the constructions can be used for giving priority to larger constructions if they unify and merge. This new reaction network is illustrated in Figure 7.3:

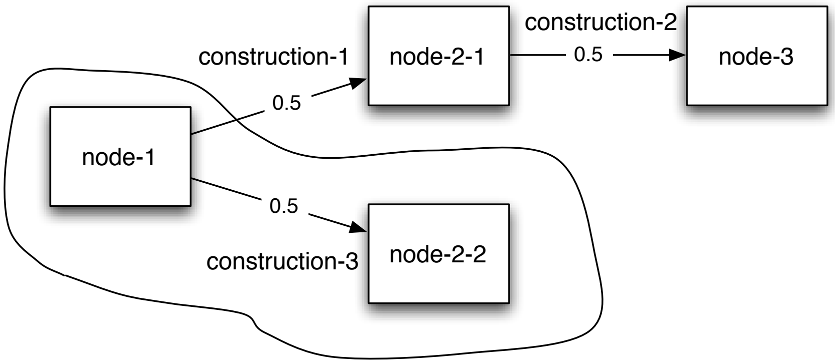


Figure 7.3: The new construction now offers the agent an alternative path in the reaction network. Since the new pattern yields the same coupled feature structure as node-3 in only one step, it is faster and therefore preferred. The links between the three constructions are used to give larger patterns priority if they unify and merge.

Apart from creating the new pattern, not much needs to be changed in the linguistic inventory apart from the fact that the agents have to link the new construction to the lexical entries that are compatible with it. The agents will not do this in one sweep but postpone this task until processing: lexical entries are only linked to the new construction instance by instance if this is required during a language game. The mechanism works entirely the same: the agent wants to unify and merge two constructions and wants to optimise processing by creating a pattern. This time, however, no new pattern needs to be created because there is already one. The pattern thus extends its use to a new verb as well. This is illustrated in Figure 7.4. The newly-made construction looks as follows:

```
<Construction:  construction-3
((?top-unit
   (sem-subunits (== ?unit-a ?unit-b ?unit-c)))
 (?unit-a
   (sem-frame (== (sem-role-1 ?unit-b ?obj-x)
                  (sem-role-2 ?unit-c ?obj-y))))
 (?unit-b
   (referent ?obj-x))
 (?unit-c
   (referent ?obj-y))
 ((J ?unit-b NIL)
   (sem-role sem-role-1))
 ((J ?unit-c NIL)
   (sem-role sem-role-2)))
<==>
((?top-unit
   (syn-subunits (== ?unit-a ?unit-b ?unit-c)))
 (?unit-a
   (syn-frame (== (syn-role-1 ?unit-b)
                  (syn-role-2 ?unit-c))))
 (?unit-b
   (syn-role syn-role-1))
 (?unit-c
   (syn-role syn-role-2)))>
```

To summarise, the agents are equipped with the following diagnostic and repair strategy in all the experiments in this chapter:

1. **Diagnostic:** If two constructions are used together for licensing a node in the network, report an opportunity for optimising processing (both for production and parsing);
2. **Repair strategy:** If there is a problem of processing effort:
   (a) If a larger construction already exists for the same mapping, create a link between the lexical entry and the construction;
   (b) Else combine the two constructions into a new construction and keep a link between them.

During processing, the link between constructions is used for giving priority to larger constructions. They can also be used for consolidation as I will show in sections 7.3 and 7.4. There are, however, no inheritance links: all relevant information is stored in the constructions themselves and no additional aspects are inherited from other constructions.

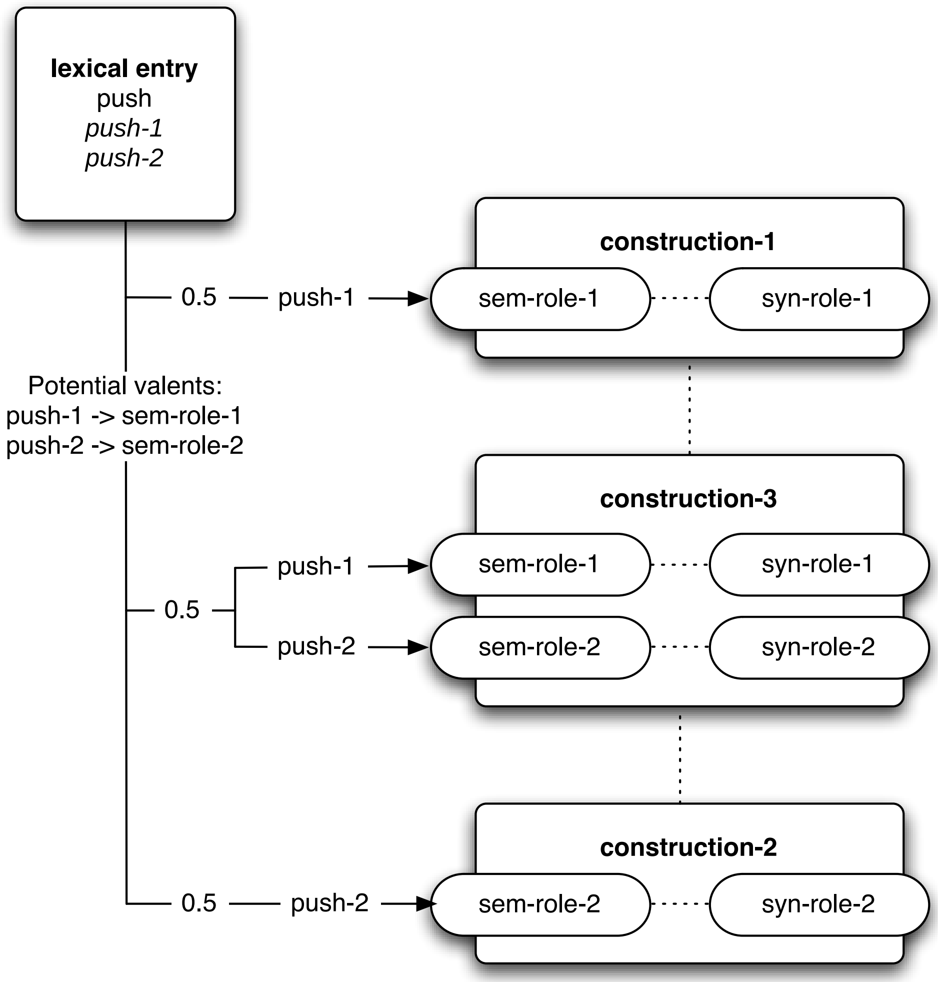


Figure 7.4: The lexical entry for *push* can now fuse with three related constructions. The lexical entry introduces the potential valents of the verb and the constructions select from these valents the actual valency. Each fusion link has its own score which can be a confidence score (between 0 and 1) or a token frequency score depending on the experiment. The three constructions were linked to each other during the creation of construction-3. This information is used during processing for giving priority to the larger construction if it unifies and merges; and in experiments 2 and 3 for the multi-level selection alignment strategy.

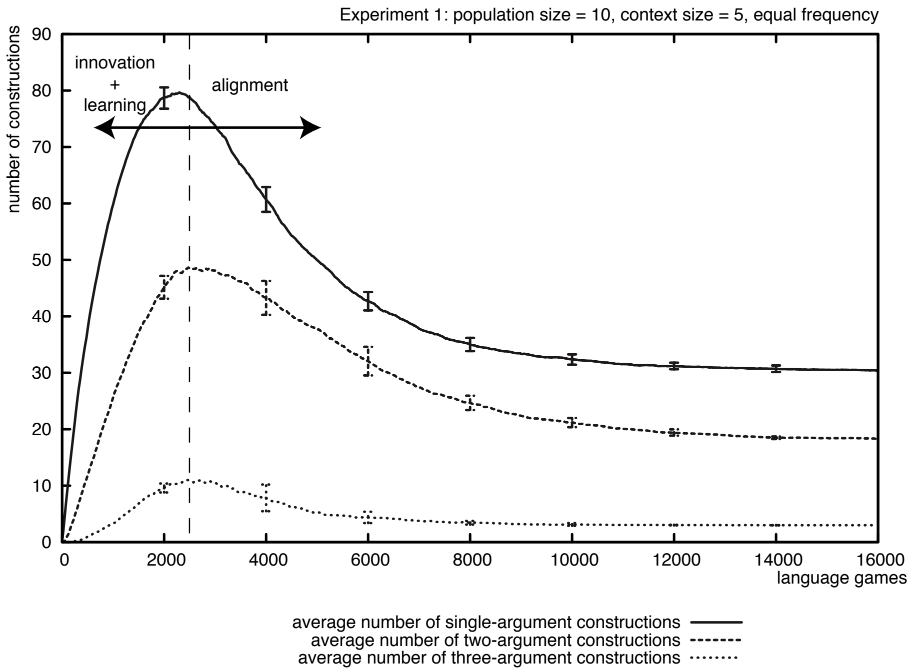


Figure 7.5: This graph shows the average number of constructions in a population of ten agents in experiment 1. In this set-up the agents succeed in converging on an optimal inventory size – given their cognitive abilities – of 30 single-argument constructions, 18 two-argument constructions and 3 three-argument constructions. The graph here indicates that there is still an average of 19 two-argument constructions but this competition also gets resolved if more language games are played.

7.2 Experiment 1: individual selection without analogy

Before immediately picking up the experiments where the previous chapter left off, the influence of the diagnostic and repair strategy for pattern formation is first tested for stage 2 in the development of case markers: the invention and adoption of specific markers.

7.2.1 Experimental set-up

The experimental set-up for experiment 1 is entirely the same as the one in baseline experiment 2c but this time the new diagnostic and repair strategy for pattern formation are added to the agents. The set-up can be briefly summarised as follows:

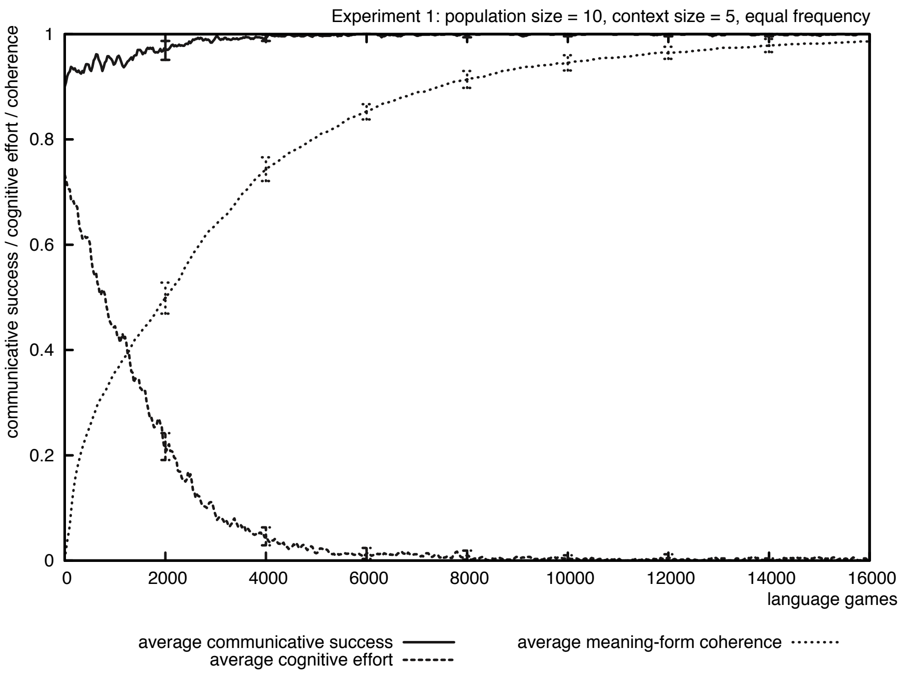


Figure 7.6: This graph shows average communicative success, cognitive effort and meaning-form coherence in a population of ten agents in experiment 1. The results show that the agents succeed in reaching 100% communicative success and reducing the cognitive effort needed for communication. Meaning-form coherence reaches almost 100% with only competition between one or two forms that is still undecided.

- The population consists of 10 agents that engage in description games;
- The meaning space is the same one as detailed in Table 6.2 and all event types occur with the same frequency;
- The agents have two diagnostics: detecting unexpressed variable equalities and the new diagnostic detecting whether two constructions were applied during processing;
- The agents have two repair strategies: one for inventing and learning new verb-specific markers and one for combining these markers into a larger construction;
- The agents use an alignment strategy of direct competition which I will further call 'individual selection'. This means that the hearer increases the confidence

scores of successfully applied constructions by 0.1 and decreases the scores of their direct competitors by 0.1. The speaker does not perform score updating.

From the above follows that the agents will have to create and converge on one construction for each possible combination of meanings. There are thirty individual participant roles that need a single-participant construction, eighteen combinations of two participant roles and three combinations of three participant roles. Since the agents have no analogy, the target number of constructions should be 51 (the sum of all these possibilities). All the combinations can be verified in Appendix A.

7.2.2 Results and discussion

Results. The experimental set-up was tested in ten series of 16.000 language games. By looking at the same measures as in the baseline experiments, the simulations seem to yield successful results at first sight. Figure 7.5 plots the average number of constructions in the population. Here, the agents have almost reached the optimal state in terms of linguistic inventory. Only in the case of two-argument constructions there are additional language games needed for deciding on the competition between one or two surviving constructions. Acquiring the constructions happens quite fast (in less than 3.000 games), but alignment takes much more time than was needed in the baseline experiments. This is due to the individual selection alignment strategy: if a pattern was used, only competing patterns are punished through lateral inhibition. The individual markers or rather the single-argument constructions they occur in are not considered during consolidation.

The long alignment period is also illustrated in Figure 7.6 which displays average communicative success, cognitive effort and meaning-form coherence. The fact that communicative success rapidly rises to 100% within 4.000 language games and that cognitive effort drops to zero between 6.000 and 8.000 language games suggests that the agents have learned all the variations floating around in their population. However, meaning-form coherence takes much longer to rise to its maximum which is again due to the alignment strategy. Coherence reaches almost 100% after 16.000 games with only competition going on for one or two cases of two-argument constructions. This competition will in the end also be resolved after additional language games.

The longer alignment period is however not the most fundamental problem with the artificial languages that are formed by the agents. A closer examination of them shows that all meaning-form mappings that they agree on are totally arbitrary. The problem is illustrated in Figures 7.7 and 7.8 which give a snapshot of convergence and coherence in one simulation after 1.000 and 7.000 language games respectively. Each meaning or combination of meanings (see Appendix A) is represented as a circle. For example, the meaning ‘approach-1’ is represented as circle 4 and meaning ‘approach-2’ is represented as circle 5. Lines between circles indicate that the meaning of one circle is a combination of the meanings of the other circles. For example, circle 31 combines ‘approach-1’ and ‘approach-2’. The colour of the circles represents the number

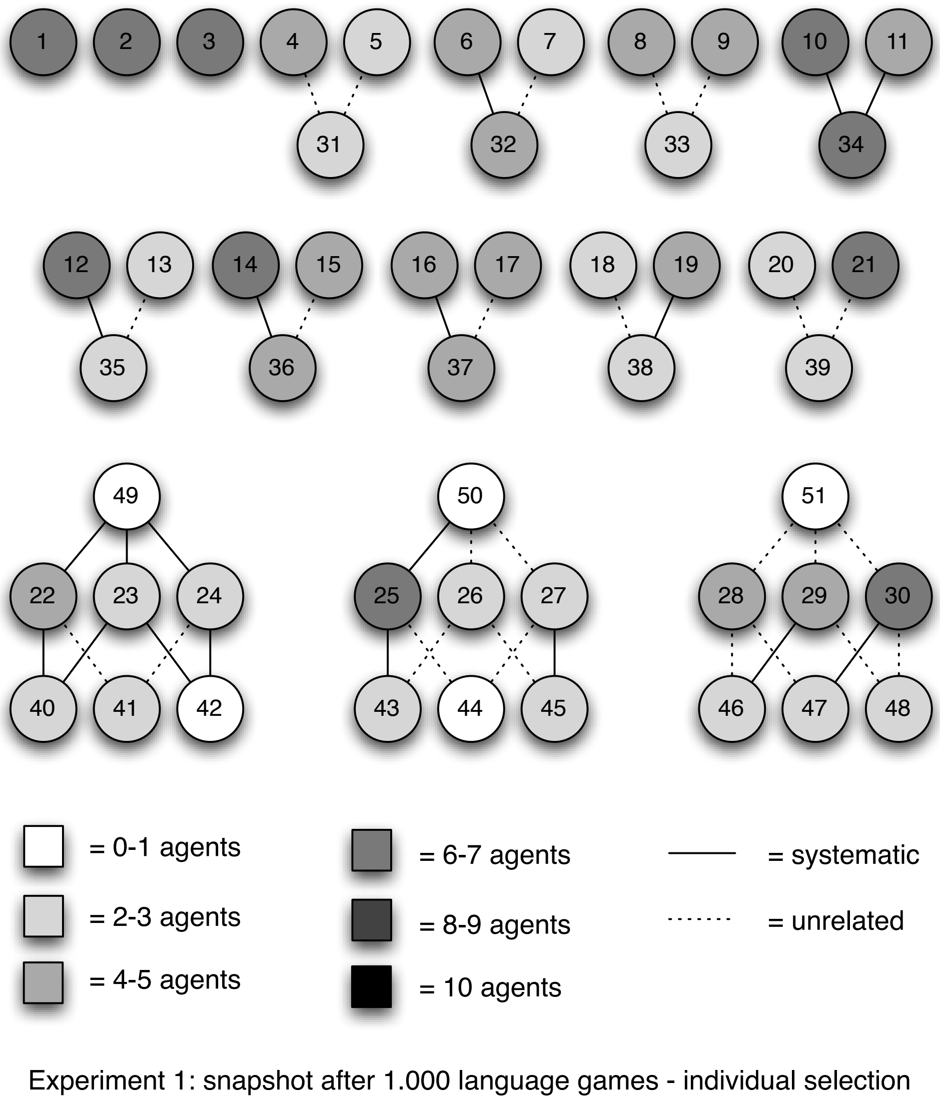


Figure 7.7: This diagram gives a snapshot of the average coherence in a population of 10 agents after 1.000 language games using the direct selection alignment strategy. Each circle stands for a particular meaning (see Appendix A), for example circles 4 and 5 stand for 'appear-1' and 'appear-2'. The lines between circles means that the meanings combine into compositional meanings, for example circle 31 means the combination 'appear-1 appear-2'. The darker the circle is coloured, the more agents prefer the same case marker(s) for covering this meaning. A full line between circles means that both meanings are covered using the same markers (= systematic), a dotted line means that a different form is preferred for the same meaning (= unrelated). The diagram shows that for most meanings only half of the population prefer the same form and that in many cases there is no systematic choice for a certain case marker.

of agents that prefer the most frequent form in the population for that particular word. A white circle means that there is either no form yet for this meaning or that there is no form which is preferred by more than one agent. A black circle means that all ten agents prefer the same form for this meaning. If all the circles are black, the agents have reached 100% convergence. If the lines between the circles are full lines, the same participant role is expressed by the same marker across constructions. If however the line is dotted, there is a different form for the same meaning.

This can best be illustrated through an example. The circle for meaning 4 (approach-1) indicates that there are 4 or 5 agents in the population which prefer the same form for marking this participant role at this stage of the simulation. For circles 5 (approach-2) and 31 (approach-1 approach-2), there are two or three agents that prefer the same form. The dotted lines between the circles, however, indicate that the most frequent pattern for circle 31 uses different markers than the single-argument constructions for circles 4 and 5:

(7.14)	*jack*	*-lich*	*approach*
	jack	approach-1	approach

‘Jack approaches (someone)’.

(7.15)	*jill*	*-sut*	*approach*
	jill	approach-2	approach

‘(Someone) approaches Jill’.

(7.16)	*jill*	*-xa*	*jack*	*-zuih*	*approach*
	jill	approach-2	jack	approach-1	approach

‘Jack approaches Jill’.

Figure 7.8 shows that after 7.000 language games, the agents have almost converged on a form for every meaning, but the problem of systematicity remains: in half of the cases, a different case marker is winning the competition on the level of single-argument constructions than the one(s) winning on the other levels. The figure also shows that in most of the cases where there is no systematic use of a form for the same meaning, convergence is also still not complete. This is in contrast to the meanings which (accidentally) arrived at the same form across constructions. Here we see mostly black circles meaning that all agents prefer the same convention.

Discussion. The results clearly indicate that the agents are not capable of constructing a systematic language. The reason for this is that all constructions are basically treated as independent linguistic items. This means that once a larger pattern is created, it starts living its own life without influencing or being influenced by the constructions that were used to create it. This results in some case markers losing the competition for marking a certain participant role on the level of single-argument constructions but still becoming the most successful one as part of a larger pattern. In all the simulations, this happened in 40 to 60% of the cases (see Figure 7.11).

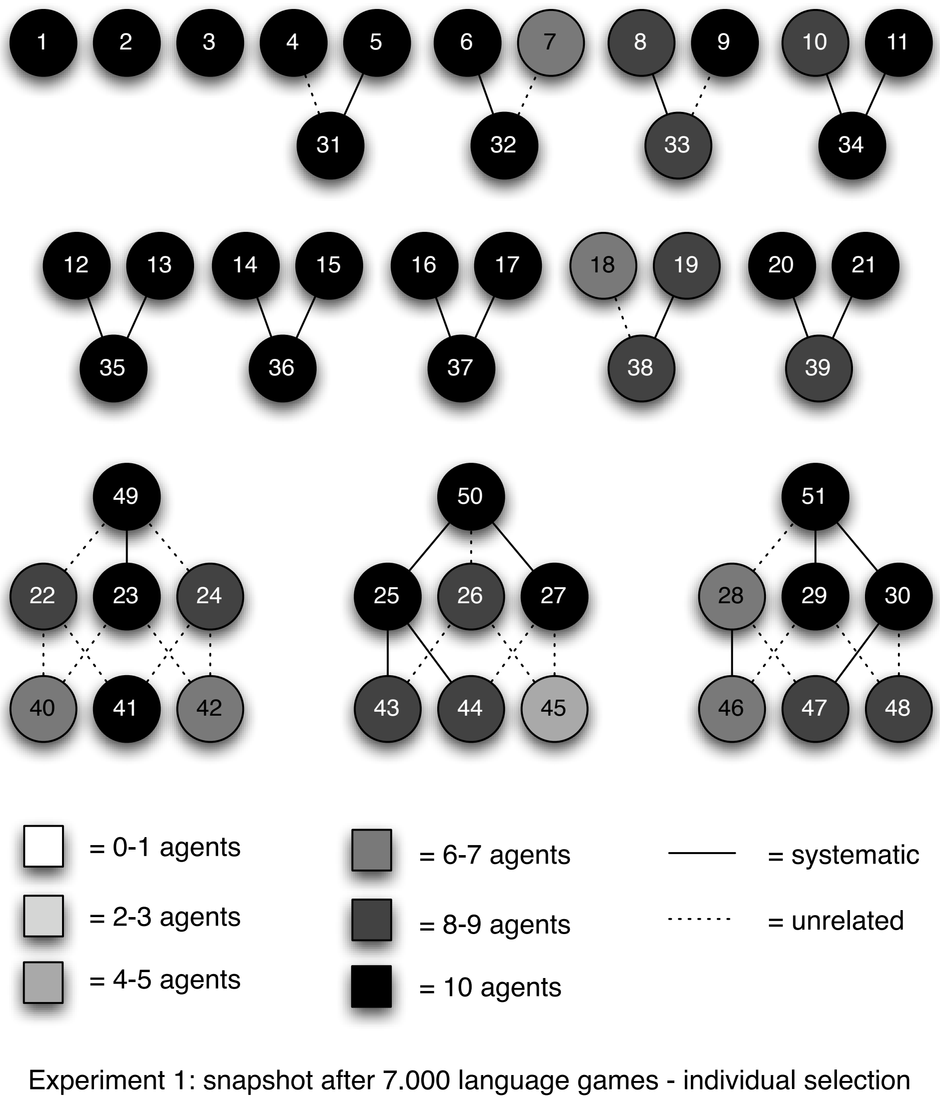


Figure 7.8: This diagram gives a snapshot of the average coherence in a population of 10 agents after 7.000 language games using the direct selection alignment strategy. The agents have reached convergence for most meanings by now, but these form-meaning mappings are not always systematically related to each other. For example, the meanings related to 49 were pretty consistent in their meaning-form mappings after 1.000 games, but have now become totally unrelated to each other: for each possible combination a new form is introduced to cover the same meaning. In all the cases where there is no systematicity, the convergence is not complete yet.

The fact that in more than half of the cases the same marker wins the competition on all levels is due to the small meaning space of the experiment and the fact that patterns are always created by combining the most successful constructions at a given point in the simulation. In fact, the agents can continue to create new patterns for a certain combination of participant roles even if they already know other patterns for it. For example, it may happen that on a lower level the average confidence scores of a new combination becomes more successful than the confidence score of the patterns. In this case the agents will still innovate which gives a slight advantage to those patterns that are in line with the most successful constructions of a lower level. As the results show, however, this is not enough.

Since natural languages are also not fully regular, it is important to see whether the lack of systematicity in the experiments is relevant for the many exceptions and sub-regularities found in natural languages. The answer is no: for most if not all irregular forms and sub-regularities in natural language, either a systematic origin can be found through diachronic changes or through external pressures such as language contact. For example, the -ed-participle in English did not manage to extend its use to all past tenses as can be observed in irregular verbs such as *to sing* and *to give*. These strong verbs are however remnants of completely regular classes of verbs in Proto-Indo-European that were able to survive thanks to their high token frequency. Despite all sociological factors, historical incidents, language contact, and other kinds of exceptions, natural languages succeed remarkably well in developing systematicity spanning over many constructions, as for example word order in English. Given the abstractions and scaffolds of the present experiments, the agents should thus be capable of developing a fully systematic language without any problems.

This leaves us the question of how systematicity can be achieved. As said before, all systematic form-meaning mappings have been formed by accident due to the small world and the nature of the innovation mechanism. For true systematicity, however, the agents need to be able to recognise relations between constructions rather than treating them as a list of independent units. This would mean that if a particular construction is successful, its systematically related constructions should also (perhaps indirectly) benefit from its success. In section 7.3 I will introduce a biologically inspired mechanism that can be exploited to achieve this effect: multi-level selection.

7.2.3 The problem of systematicity in other work

As to my knowledge, the problem of systematicity has never been reported before in the field of the origins and evolution of language. This does not mean, however, that the problem never existed. In this section, I will give a brief overview of some prior work in the field in which the problem was either overlooked or in which it could not occur due to experimental assumptions.

Exemplar-based simulations. One computational simulation which is closely related to the work in this thesis is presented by Batali (2002). Batali investigates how a multi-agent population can form a recursive communication system by using exemplars stored in memory. This work can be categorised as a 'problem-solving model' (see section 3.1.3) because these exemplars have to be agreed upon in locally situated interactions. Each exemplar has a confidence score which is increased and decreased according to similar lateral inhibition dynamics as in the simulations of the previous section. The type of learner is thus the same one as the agents in this thesis: they build their language instance by instance in a bottom-up and redundant fashion. Batali's agents only keep exemplars and all generalisation in the model is captured by directly manipulating these exemplars during processing. Here is an example of an exemplar composed of two smaller ones (Batali, 2002, exemplar 5.1.2.a):

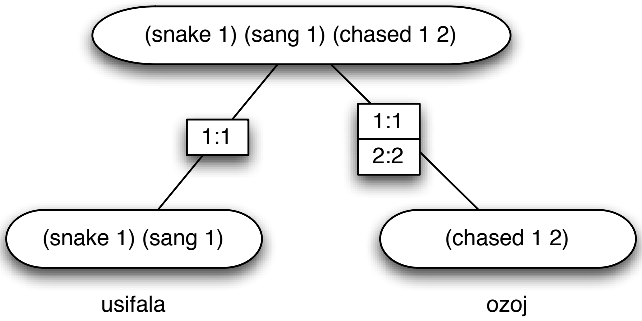


Figure 7.9: A complex exemplar from Batali (2002). The exemplar features a compositional meaning with 'argument maps' to the smaller exemplars that take care of variable equalities in the meaning.

Batali does not use event-specific variables as I do in this thesis but assumes a simple three-way contrast between arguments 1, 2 and 3. For example the meaning ((snake 1) (sang 1)) translates to something like 'the snake sang', whereas ((snake 1) (sang 2)) would mean something like 'there was a snake and something sang.' Event structure is stored immediately in the exemplar but can be overridden by argument maps between complex exemplars and their subcomponents. For example, the argument map '1:2' translates a meaning like (rat 1) to (rat 2). These argument maps are also stored as part of the complex exemplar. Apart from these argument maps between complex exemplars and their components, all exemplars are unrelated and listed in the memory.

The agents then engage in a series of description games. They are able to invent new words for new meanings and they are capable of combining existing words into larger patterns or breaking up a pattern again into smaller parts. The ultimate goal of the agents is two-fold: (a) agree on a shared lexicon for all the single meanings (e.g. cat, fox, chase, etc.) and (b) agree on a way to mark event structure through the argument maps (i.e. marking the difference between arguments 1, 2 and 3). The simulations make use of a single generation of agents.

The results indicate that the agents gradually reach communicative success and that they agree on the same exemplars. Goal (a) is therefore definitely reached. However, the results show that event structure is not always marked in the same way: all the simulations end up using specific ordering for each exemplar (even though they may involve the same meanings) and using 'empty' words that accidentally evolved into markers for argument mappings. The agents thus do not succeed in agreeing on a systematic way of distinguishing participant '1' from participants '2' and '3'. The agents thus cannot generalise argument mapping to new predicates such as (give 1 2 3) and have to negotiate event structure for each word separately. This lack of systematicity is not noted by Batali as a problem and the use of the empty words is wrongly interpreted as corresponding to argument markers in natural languages.

Probabilistic grammars. Another experiment in which the systematicity problem is overlooked is reported by De Pauw (2002, chapter 10). De Pauw investigates how rudimentary principles of syntax can emerge from distributional aspects of communication rather than from the interface between syntax and semantics. This exclusive focus on syntax is different from the work in this thesis (even though some semantics is smuggled into De Pauw's simulations in the disctinction between animate and non-animate objects which results in different distributional patterns). Similar assumptions to this thesis are the heavy use of memory (even more so by De Pauw), a bottom-up and redundant formation of the language and a predefined lexicon in order to focus exclusively on the topic of interest. The population in De Pauw's simulations is dynamic in the sense that there is a generational turn-over, but no linguistic information is transmitted genetically from one generation to the next.

The agents engage in a series of language games in which they communicate about objects or events. If there are several objects, the agents can choose between six word orders: SVO, SOV, VSO, VOS, OSV or OVS. De Pauw therefore does not distinguish between verb-specific participant roles, but only assumes a two-way contrast between the subject (S) and the object (O). The agents start without any preference for a particular word order so variation naturally occurs in the population. The alignment strategy of the agents is simply storing bigrams or frequencies of co-occurrences and performing statistical induction on top of those bigrams (De Pauw, 2002, p. 362):

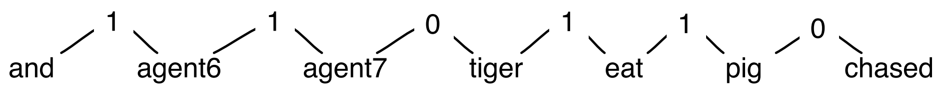


Figure 7.10: The agents in De Pauw (2002) store co-occurrence frequencies and use these bigram-probabilities to decide on a preferred ordering.

During the simulations, the agents rapidly converge on fixed word order on simple relations. However, as De Pauw notes, there are no general tendencies in terms of a general word order. Each 'verb' rather has its own preferred ordering. For more complex relations, the agents do not always reach coherence. De Pauw concludes that the agents therefore reside in a local maximum and that they evolve from one local maximum of convergence to the next. De Pauw argues that this is not a shortcoming of the model but rather its greatest asset: whereas other models in the field are looking for the state of convergence, *"language itself never converges and constantly adapts to a changing environment and seems to be driven by chaotic elements, introducing a large degree of randomness in language both from a synchronic, as well as a diachronic point of view"* (p. 378).

There are however no such chaotic elements present in De Pauw's model which should prevent the agents from reaching complete coherence. The degree of randomness in his simulations seems to stem from the systematicity problem: by only looking at bigram probabilities, it is to be expected that there is an arbitrary word order which is verb-specific. In the case of more complex predicates, the preferred order depends on a combination of various bigrams which increases the randomness because the probabilities of these bigrams are constantly changing so it becomes much harder to agree on a fixed order for these complex meanings. De Pauw dismisses the possibility that the degree of convergence is the maximum that can be expected from the population, but this is in fact the only correct conclusion. Given the cognitive capabilities of the agents, convergence could only increase if the input would be more structured. In certain machine learning tasks, there is already a lot of structure present in the learning data so bigrams can be successfully used for making some predictions. In the case of language formation, however, agents have to start from scratch so there is no structure spanning multiple levels yet that can be induced.

De Pauw's concluding remark is that it is empirically impossible to know whether the agents succeed in *"expressing the proper (agent, patient) relationship, or if it is just a side-effect of beneficial bigram probability distributions"* (p. 376). I would argue, however, that since the agents are not endowed with the capacity of relating bigrams to each other but solely rely on these probabilities, all tendencies in word order are in fact a side-effect of the bigrams. The conclusion is that De Pauw (2002), just like Batali (2002), misinterpreted experimental results because the problem of systematicity was not noticed.

Iterated Learning Models. So far I only discussed models that featured lazy learners: agents which postpone generalisation until processing time and which shape their language in a step-by-step fashion. As opposed to lazy learners there are 'eager learners'. Eager learners try to look for generalisations (and abstractions) before it is actually needed in processing and work on the complete inventory. Eager learners typically discard the examples that can be derived from a rule and thus try to optimise the inventory size. If the problem of systematicity also occurs with eager learners, then we know that the problem is not exclusive to the usage-based approach proposed in this thesis.

In the field of artificial language evolution, especially Iterated Learning Models (see section 3.1.2) feature agents that loop through their inventory after each interaction in order to make abstractions. In section 8.1 I will draw a thorough comparison between my experimental results and those of Moy (2006), who investigated the same topic using the Iterated Learning Model so I will not go into details here. As a quick preview, I can already give away one of the conclusions which is that the problem of systematicity also occurs in Iterated Learning Models. This does not only happen in Moy's experiments, but also in the simulations reported by Kirby (2000) and Smith *et al.* (2003) even though these models feature complete meaning transfer and a population of only two agents.

The conclusion is the same as for the other simulations reported in this section: the problem of systematicity goes by unnoticed in most Iterated Learning Models, but becomes very apparent in Moy (2006). The problem occurs for the same reasons as in all the other experiments: the agents only behave 'systematic' during innovation and learning, but then treat all linguistic items as an unstructured list of unrelated elements. So either there is no adequate model yet that avoids the problem of systematicity or the problem is **not restricted to the type of learner**. In the latter case, the problem seems to be caused by the fact that the linguistic inventory is unstructured.

Other models. Finally, there are many models that investigate certain aspects of grammar in which the problem of systematicity does not occur such as De Beule (2007); De Beule & Bergen (2006); Nowak & Krakauer (1999); Steels & Wellens (2006); etc. I will take the simulations by De Beule & Bergen (2006) as an example of why these models don't have the problem. The conclusions of this brief discussion extend to all the other models on grammar as well.

De Beule & Bergen investigate the competition between holistic and compositional utterances (see section 3.1.3). In case of compositional utterances, one could expect the problem of systematicity to pop up, but it doesn't. The reason is that De Beule & Bergen designed their experiment in such a way that the agents had prior knowledge about what kind of categories and constructions to expect: individual words are immediately tagged with a certain syntactic category and grammatical constructions are fully schematic from the start. One construction can hence be used for all possible combinations of competing individual words that are tagged with the same category and remains agnostic as to which words should win the competition. The experiment thus made a clear separation between the lexicon on the one hand and grammatical constructions on the other; and it did not offer the agents the possibility of in-between patterns or idioms.

This is not a criticism of the model per se: given the fact that De Beule & Bergen only intended to focus on competition between holistic and compositional utterances, the design choice is justified in which the competition dynamics can be clearly investigated on each level. As such the experiment can be interpreted as investigating a prerequisite of grammar rather than the formation of *actual* grammar. For the scope of this thesis, this experimental design is thus not warranted: the barrier between fully idiomatic items and fully schematic items needs to be broken down.

7.2.4 Conclusions

In this section, I reported the results of an experiment in which a population of ten agents were able to invent and learn verb-specific markers and to combine these markers into larger patterns. The results showed that the agents succeeded in converging on an optimal inventory size for the task at hand. However, a closer look at the formed languages showed that there was no systematicity: constructions that win the competition on a higher level can contain different forms from the ones used by constructions that won the competition on a lower level.

In a brief overview of related experiments in the field, I showed that the problem of systematicity is neither exclusive to the set-up of experiment 1, nor to the type of learner proposed here. The problem in fact automatically occurs if the linguistic inventory is an unstructured list of items. Experiments on grammar that did not have the problem of systematicity have built some kind of structure in, such as a sharp distinction between lexical items and grammatical constructions. The problem of systematicity has remained unnoticed in the field so far and has led to wrong interpretation of experimental results. In all cases where the problem occurs, the experimenter maps the irregularities or lack of complete convergence to irregular forms in natural languages. I argued that this mapping is wrong since the causes of the irregularities are completely different. Given the abstractions, scaffolds and assumptions of these computational simulations, systematic coherence and convergence should be possible.

7.3 Experiment 2: multi-level selection without analogy

In the previous section I demonstrated the problem of systematicity that occurs during the formation of a language if the agents treat all entries in their linguistic inventory as unrelated individuals and if their language comprises multiple layers of organisation. An alignment strategy involving the individual selection of constructions leads to completely arbitrary form-meaning pairs whereas natural languages show greater cohesion and a higher degree of systematically related constructions. Even in idioms such as *he kicked the bucket*, some degree of schematicity is present such as the conjugation of the verb. The agents therefore need a new alignment strategy in which the success of one construction may have an impact on the success of other related constructions.

In this section I will present an experiment which features new alignment strategies which are inspired by the notion of 'multi-level selection' in evolutionary biology (Wilson & Sober, 1994). Multi-level selection (formerly known as 'group selection') acknowledges the fact that groups or other higher-level entities can act as 'vehicles' for selection. In this view, not all aspects of groups are reduced to by-products of individual (and usually selfish) interactions. In other words, being part of a group can increase the selectionist advantage of individuals.

Natural languages are clear instances of organisms with a hierarchical functional organisation which can be conceived as 'groups within groups'. Competition is going on at multiple levels of this organisation: between synonyms for becoming dominant in expressing a particular meaning, between idiomatic patterns that group a number of words, between different syntactic and semantic categories competing for a role in the grammar, between ways in which a syntactic category is marked, etc. Multi-level selections therefore seems to be readily applicable to language as well.

7.3.1 Experimental set-up

The most important requirement for implementing multi-level selection is that the agents have to be capable themselves of recognising relations between linguistic items. This is in fact not so difficult to achieve: in section 7.1.2 I explained that the agents keep a link between larger constructions and the constructions that were used for creating them. These links can now be used for implementing multi-level selection. Three different alignment strategies have been implemented for comparison:

- **Top-down selection**: if the game was a success, the hearer will not only reward the constructions that were applied during processing, but also all the related constructions on a lower level. The confidence scores of all the competitors of these constructions are decreased through lateral inhibition.
- **Bottom-up selection**: If the game was a success, the hearer will not only increase the score of the applied constructions, but also the scores of all the related constructions on a higher level. All the competing constructions are punished.
- **Multi-level selection**: If the game was a success, the hearer will not only increase the score of the applied constructions, but also the scores of all related constructions. All the competing constructions are punished through lateral inhibition.

Retrieving related constructions is performed recursively. For example, if a three-argument construction was applied using the top-down selection alignment strategy, its two sub-components are retrieved (a two-argument and a single-argument construction) as well as the two sub-components of the two-argument construction. The hearer thus increases the scores of five constructions. The competitors are all the direct competitors of these five constructions.

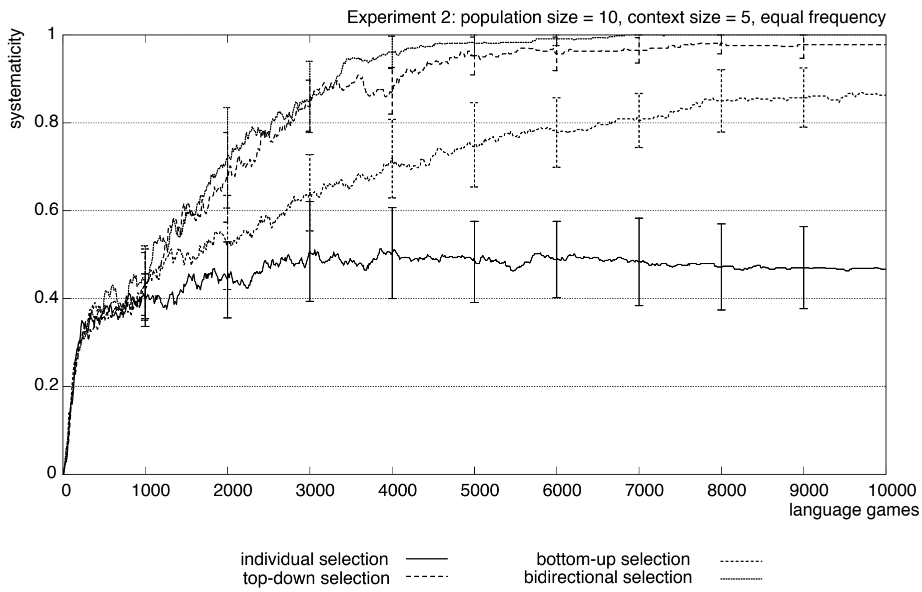


Figure 7.11: This graph compares the performance in terms of systematicity of four experimental set-ups. systematicity fluctuates around 50% in the baseline case where there is only individual selection (experiment 1). With the bottom-up selection strategy the agents improve the systematicity rate to 80% but then get stuck. Top-down selection leads to full systematicity in some of the runs, but most of the simulations feature some 'frozen accidents' as well. Only the multi-level selection strategy leads to full systematicity in all the series after about 7.000 language games.

During processing, only the scores of the applied constructions are taken into account and not of the whole group of related constructions. The group selection dynamics therefore only matter during consolidation. The rest of the set-up is the same as for experiment 1.

7.3.2 Results and discussion

The three alignment strategies were compared to each other and to experiment 1 in ten series of 10.000 language games. Special attention was given to the measures 'meaning-form coherence' and 'systematicity' (see Appendix A).

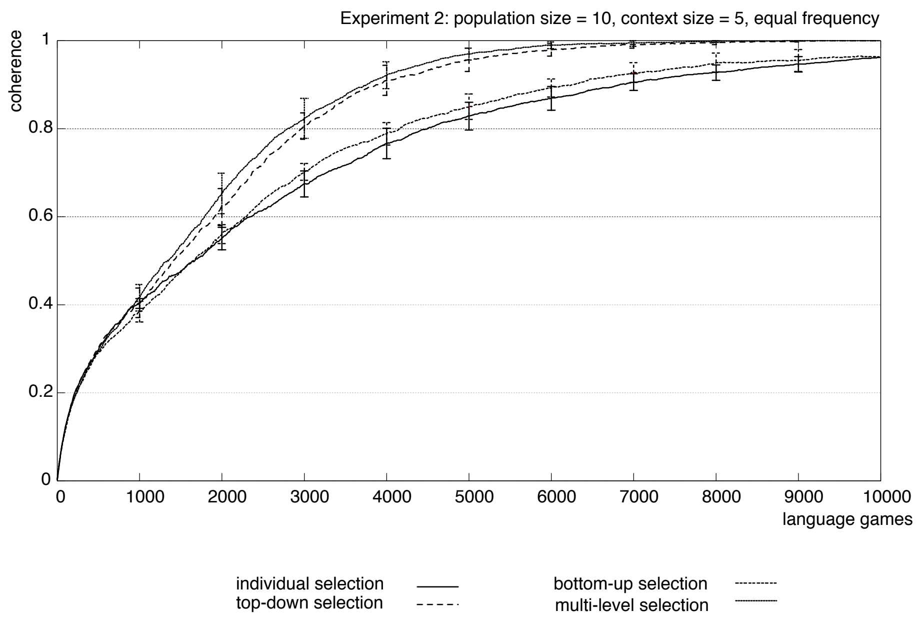


Figure 7.12: Since the systematicity graph only takes the most frequent forms into account, meaning-form coherence has to be checked in order to verify whether all the agents have converged on the same form-meaning pairs. We see that after 10.000 language games, only the top-down and the multi-level selection alignment strategies have already reached complete coherence. Multi-level selection slightly outperforms top-down selection but not significantly so. In the case of bottom-up and individual selection, the agents need additional language games for reaching coherence.

Results. Figure 7.11 illustrates the amount of systematicity in all four alignment strategies. The graph shows that the three alignment strategies involving multiple levels all improve on the baseline of individual selection of experiment 1. With the alignment strategy of individual selection, systematicity fluctuates between 40 and 60% depending on how 'lucky' the agents were. The behaviour of the other three strategies is much more consistent over the ten series. The graph shows that bottom-up selection allows the agents to improve systematicity to 80% but there they are faced with 'frozen accidents' as well. The top-down selection improves systematicity even further and allows the agents to reach full systematicity in some of the runs. However, in most cases, there were still two or three unsystematic patterns left. Only the multi-level selection strategy led to full systematicity in all the simulations.

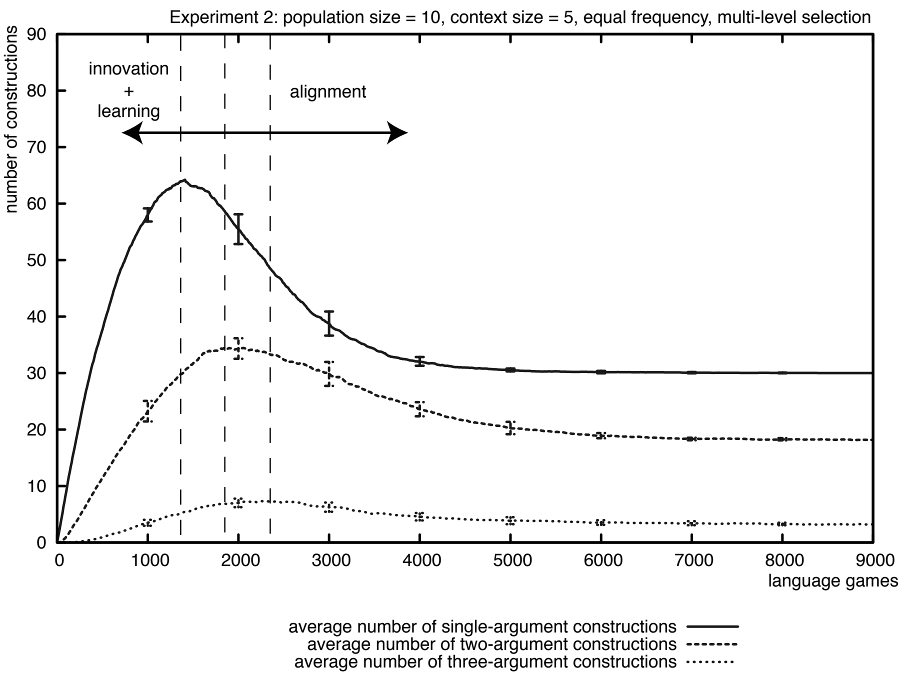


Figure 7.13: This graph shows the average number of constructions known by an agent using the multi-level selection alignment strategy. Compared to individual selection, multi-level selection allows the agents to discard competitors much more rapidly: there are significantly less variations floating around in the population. For example, the peak of single-argument constructions is about 60 instead of 80 in experiment 1. Also the alignment phase happens much faster.

Since the measure of systematicity only looks at the most frequent forms floating in a population, it needs to be complemented with meaning-form coherence to verify whether *all* the agents converge on the same preferences. Figure 7.12 therefore compares the performance of the four alignment strategies in terms of coherence. From the results of experiment 1 we already knew that in the case of individual selection, alignment takes longer than 10.000 language games. The coherence line for bottom-up selection runs almost parallel with it and does not improve on it in terms of convergence speed. The only two strategies that reach convergence within 10.000 games are top-down and multi-level selection. Full coherence in the case of top-down selection however does not mean full systematicity, as was shown in Figure 7.11.

Figure 7.13 shows the average number of constructions in the ten series involving the alignment strategy of multi-level selection. The graph confirms the fact that the agents converge significantly faster on an optimal number of constructions than in experiment 1. At the peak of competing constructions, there are about 60 single-argument constructions, 30 two-argument constructions and 6 three-argument constructions or an average of two competing constructions for each possible meaning. This is much less than in experiment 1 (see Figure 7.5) which featured peaks of 80 single-argument, 50 two-argument and 10 three-argument constructions. Also alignment happens much faster: with multi-level selection, the agents align after 6.000 games as opposed to 14.000 language games or more if the agents use individual selection.

Figures 7.14 and 7.15 offer a snapshot of the most frequent forms in a population using the multi-level selection alignment strategy. Both snapshots confirm the results indicated by the coherence and systematicity graphs. Figure 7.14 shows already much more dark grey circles than Figure 7.7 featuring individual selection indicating that for most meanings there is already a majority of agents preferring the same form. The preferred forms are also to a higher degree systematically related to each other than in experiment 1, even for the more complex patterns. All black circles in the Figure feature meanings which are related to other meanings which suggests that multi-level selection indeed favours groups of related items. The snapshot in Figure 7.15 only shows black circles which means that all agents in the population prefer the same form for that particular meaning. There are also only full lines between the circles indicating that the same case markers are consistently used across patterns. This result significantly improves over the results of the simulations in which agents used the strategy of individual selection.

Discussion. The results show that the agents reach full systematicity if each level in the linguistic inventory can have an influence on the competition in other levels. It is now important to understand why this is the case and why the other alignment strategies did not yield full systematicity.

First of all, bottom-up selection doesn't improve on the results of experiment 1 in terms of convergence speed and it doesn't lead to a systematic mapping of meaning to form across patterns. A closer examination of the the alignment strategy reveals that the relatively high systematicity of 80% is due to the higher frequency of single-argument constructions. The frequent use of these constructions each time has repercussions for the competition between larger constructions whereas this is not the case in the other direction. If there would be more patterns than single-argument constructions, the improvement would therefore be less high. The patterns that resist the bottom-up selection can do so because the competition is not fully decided yet on a lower level (each time reinforcing competing patterns) and there is additional competition on the level of the patterns themselves which may have different winners than those on lower levels. In the end, though, the bottom-up strategy should lead to full systematicity but only very slowly and because the smaller constructions are more frequent.

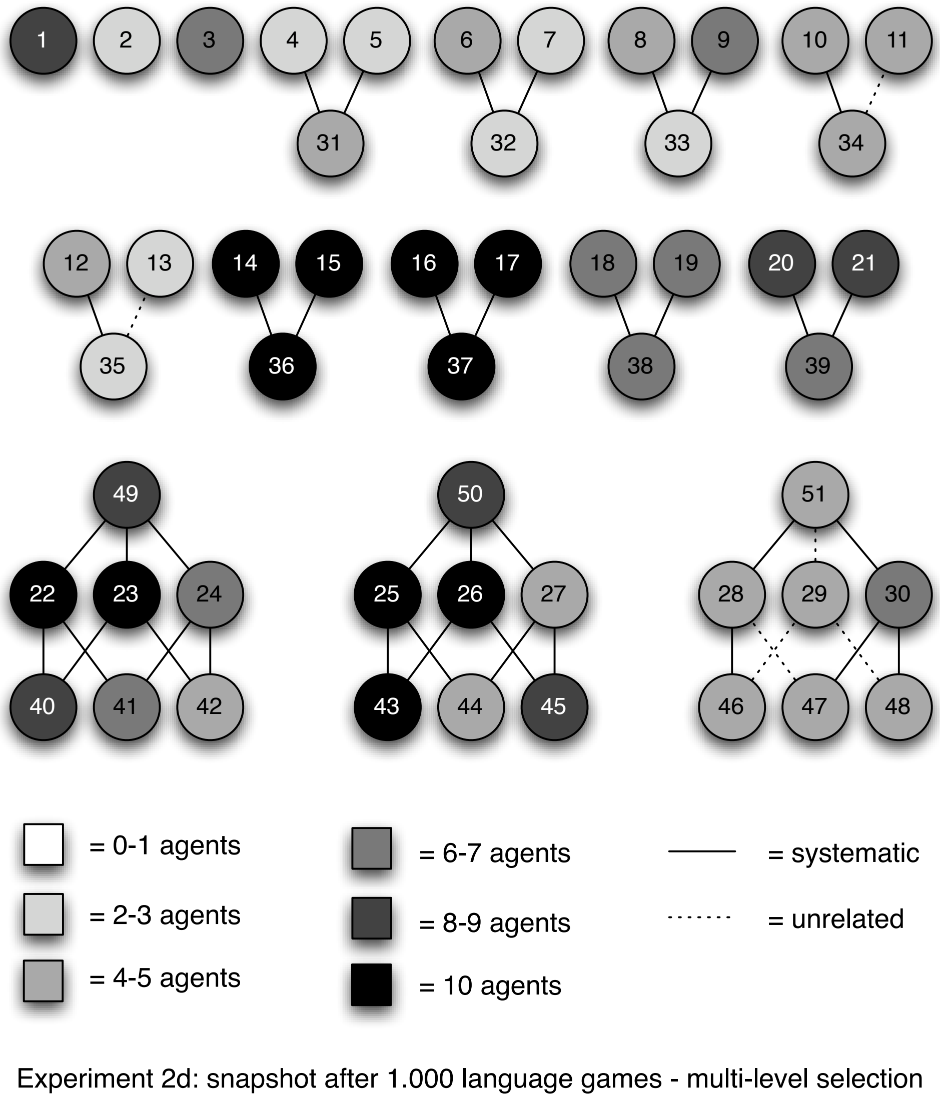
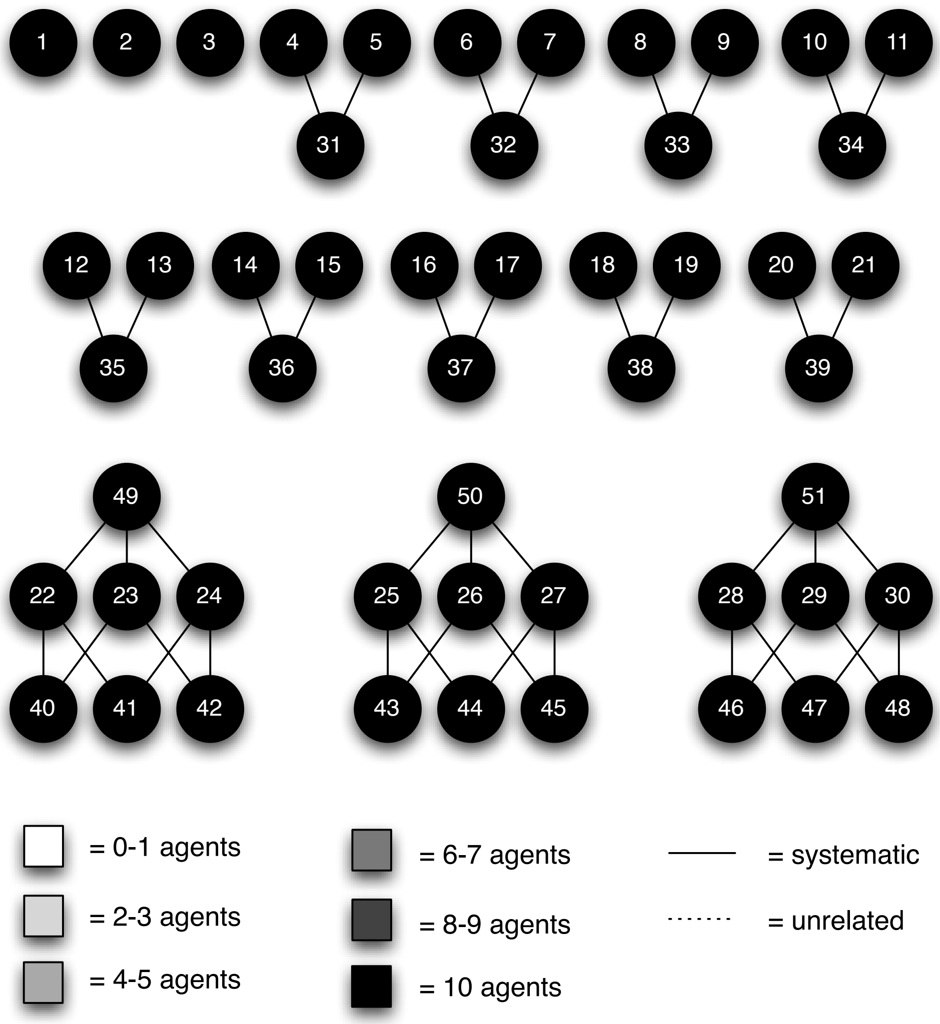


Figure 7.14: This diagram gives a snapshot of the average coherence in a population of 10 agents after 1.000 language games using the multi-level selection alignment strategy. Compared to the simulations using direct selection, the agents seem to be converging more rapidly on meaning-form mappings and have already settled on 11 of them. Whereas there was no convergence at all yet for the more complex meanings 49–51 in the simulations using the direct selection alignment strategy, here they are already shared by a majority of the population. This suggests that multi-level selection speeds up the convergence dynamics significantly especially for related meanings. For the meanings in the bottom right, there is less systematicity and hence convergence takes longer time.



Experiment 2d: snapshot after 7.000 language games - multi-level selection

Figure 7.15: This diagram gives a snapshot of the average coherence in a population of 10 agents after 7.000 language games using the multi-level selection alignment strategy. All the circles are black which means that the entire population prefers the same meaning-form mappings. The language is also fully systematic so the agents have converged on 30 case markers which can be combined into 51 different constructions. The results indicate that even for instance-based learners/innovators systematicity can be reached by keeping a link between newly acquired constructions and the constructions that were used for learning or creating them.

The top-down strategy is affected by frequency as well: competition between the larger constructions has significant impact on lower levels because it can increase the scores of up to six constructions while at the same time punishing all competitors. However, since the smaller constructions are actually more frequent than the larger ones, some divergent competition pathways may resist this influence from the patterns and survive nevertheless. As the results indicate, this in fact happens in most of the simulations. Top-down selection thus improves systematicity significantly, but it is affected by the frequency of the various levels of linguistic items and it is therefore no guarantee of full systematicity.

Finally, the agents can achieve full systematicity through multi-level selection. This strategy allows the competition of each level to influence the competition on others and given its n-directionality, it is not (or less) dependent on differences in frequency. Moreover, the agents do not need to differentiate between a ‘higher’ and a ‘lower’ level but can treat all links between constructions on equal footing. The results of experiment 2 confirm earlier results on multi-level selection and systematicity reported by Steels *et al.* (2007). In these experiments, which involved a scale-up in convergence space, multi-level selection outperforms the other strategies even more significantly.

The exact nature of multi-level selection and how it relates to multi-level selection in biological and social phenomena (among which natural languages) is still unclear and scaling up the experiments in terms of population and meaning size may require a further refinement of the algorithm. Especially when there can be interference between competing systems the algorithm probably needs some rethinking. However, the comparison between experiment 1 and 2 clearly demonstrate that it is an indispensable mechanism for keeping an evolving language systematic.

7.3.3 Conclusions

Experiment 2 demonstrated the effect of various alignment strategies on the convergence speed and systematicity of the artificial languages developed by the agents. The results indicate that multi-level selection is a crucial mechanism for keeping languages coherent and preventing them from becoming a collection of unrelated and arbitrary form-meaning mappings.

Based on the experimental results and prior work on multi-level selection and pattern formation, I argued that the best alignment strategy is the one where not only components can profit from the success of larger patterns, but where larger patterns are also influenced by evolutionary processes on a lower level. This strategy also has the advantage that the agents do not make an actual distinction between lower and higher levels, but can treat all related constructions in the same way.

7.4 Experiment 3: multi-level selection with analogy

Similarly to the previous experiments, experiment 3 investigates how baseline experiment 3 can be extended with a diagnostic and repair for pattern formation. Since the previous experiments identified the problem of systematicity, experiment 3 first of all needs to verify whether the conclusions of the second experiment also hold for the new set-up in which the agents are capable of performing analogical reasoning over events. I will then report an experiment that adapts the algorithm for multi-level selection to the token-frequency alignment strategy of baseline experiment 3d (see section 6.4).

7.4.1 Experimental set-up

Experiment 3 features the same experimental set-up as baseline experiment 3 with the addition of a diagnostic and repair strategy for pattern formation. To summarise:

- The population consists of 10 agents that engage in description games;
- The meaning space is the same one as detailed in Table 6.2 and all event types occur with the same frequency;
- The agents have two diagnostics: detecting unexpressed variable equalities and the new diagnostic detecting whether two constructions were applied during processing;
- The agents have two repair strategies: one for inventing and learning new verb-specific markers and one for combining these markers into a larger construction. The invention and learning strategy also includes the possibility of extending and reusing existing markers through analogical reasoning. The algorithm for analogy is the same as in baseline experiment 3 and only looks at individual markers.

The experiment has been tested using five different alignment strategies. The first four strategies are individual selection, top-down selection, bottom-up selection and multi-level selection using the same fine-grained lateral inhibition mechanism as used in baseline experiment 3c. This means that competition is only held at the level of the co-occurrence links between a construction and a lexical entry rather than at the level of the constructions themselves. The algorithms can be summarised as follows:

- **Individual selection:** This is the exact same set-up as baseline experiment 3c. If a game was successful, the hearer will increase the score of the co-occurrence link between the applied lexical entry and the applied construction(s) by 0.1. She will also decrease the scores of the competing links by 0.1. The score of a link is always between 0 (high uncertainty) and 1 (high confidence).
- **Top-down selection:** In this strategy, the hearer will not only increase the score of the relevant co-occurrence link, but also the score of all the co-occurrence

links that link the lexical entry to the smaller constructions which are related to the applied construction. The scores of the competitors of these links are decreased.

- **Bottom-up selection:** In this strategy, the hearer increases the score of the relevant link and of all the co-occurrence links that link the lexical entry to the larger constructions which are related to the applied constructions. All competitors of these links are punished.
- **Multi-level selection:** In this strategy, the hearer increases the scores of the relevant co-occurrence links and of all the links which link the lexical entry to constructions that are related to the applied constructions.

The fifth experimental set-up does not involve lateral inhibition but implements **multi-level selection and memory decay**. In this set-up, the hearer will not only increase the frequency score of the applied constructions, but also that of all the related constructions by 1. The frequency scores have no upper limit, so the higher the score, the more entrenched the construction is. After an agent has individually engaged in 200 language games, the frequency scores of all the items in the inventory are decreased.

In all five set-ups, the speaker will use the co-occurrence links to speed up processing. This means that not the entire inventory of constructions is considered, but only those constructions which are linked to the lexical entry. Links can be added through co-occurrence. When the speaker is faced with multiple hypotheses, she will choose the construction which either has the strongest co-occurrence link with the lexical entry (in the first four set-ups) or the one with the highest token frequency (in the fifth set-up). During processing, only the scores of individual competitors are taken into account. All simulations have been run in 10 series of 12.000 language games.

7.4.2 Results and discussion

In this section, the first four set-ups are again compared to each other to demonstrate the reoccurrence of the problem of systematicity. The fourth set-up (multi-level selection with lateral inhibition) is then compared more thoroughly to the fifth set-up using multi-level selection and memory decay. Finally, this section offers a closer look at one language evolved using the fifth set-up.

Results. Figure 7.16 compares the first four experimental set-ups to each other in terms of systematicity. The lines indicating systematicity for each set-up show the same behaviour as those in experiment 2 (Figure 7.11). The first set-up using the alignment strategy of individual selection fluctuates between 45 and 60% systematicity and stops evolving after 6.000 language games. The bottom-up strategy reaches more than 80% systematicity after 8.000 language games. In some simulations, this strategy leads up to 90% but never to maximum systematicity. Top-down selection performs a bit better than in experiment 2 due to the fact that the agent's capacity of reusing

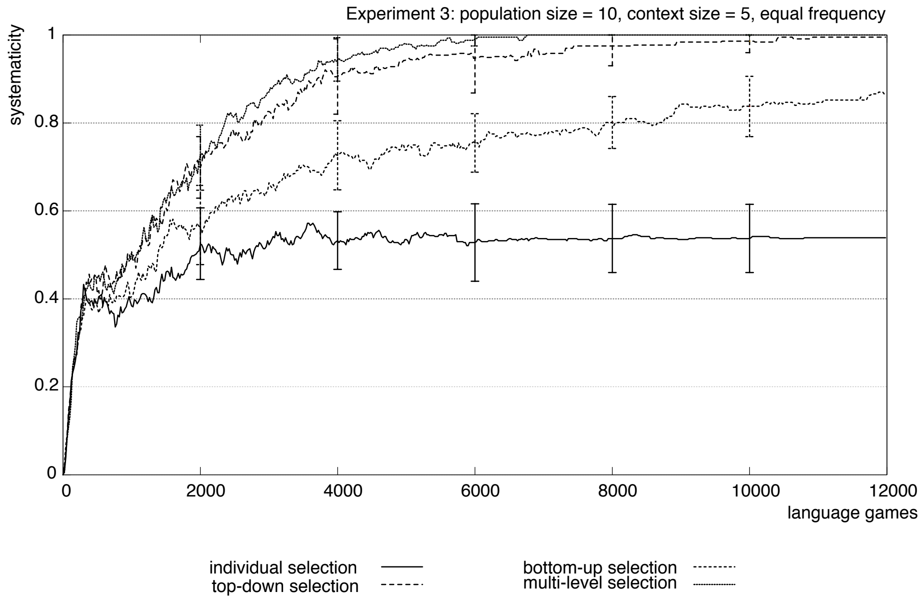


Figure 7.16: This graph compares systematicity in the first four experimental set-ups of experiment 3. The results confirm those of experiment 2. Even though there is potentially less variation because of the reuse of existing markers, individual selection stagnates at 50% systematicity. Bottom-up selection increases systematicity beyond 80% but also gets stuck. Top-down selection manages to reach full systematicity in more simulations than in experiment 2 because of the smaller variation space, but does not guarantee full systematicity. Only multi-level selection reaches systematicity in all the simulations and does so significantly faster than the other alignment strategies.

existing markers leads to a smaller variation space so 'frozen accidents' are less likely. Yet, as the results show, some simulations still involve an unsystematic convention and reaching systematicity takes a longer time than the multi-level selection strategy. The latter strategy is again the only one which leads to full systematicity in all the simulations.

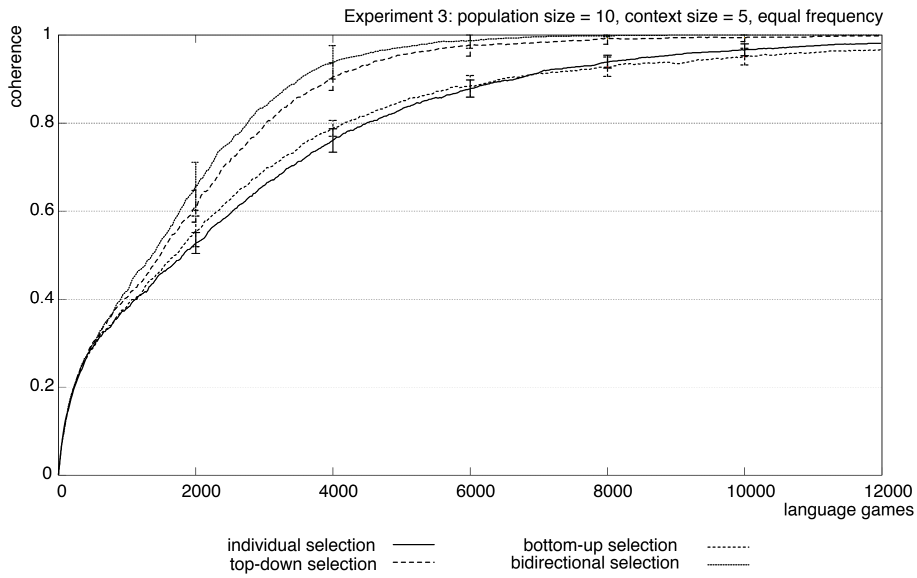


Figure 7.17: This graph compares the first four set-ups of experiment 3 in terms of meaning-form coherence. The graph shows that multi-level and top-down selection perform equally well in terms of coherence and reach 100% between 6.000 and 8.000 language games. Bottom-up and individual selection again run almost parallel and reach coherence faster than in experiment 2 because of the smaller variation space.

The four set-ups also confirm the results of experiment 2 in terms of coherence. Figure 7.17 shows that bottom-up selection and individual selection again run almost parallel in terms of convergence. This time the agents reach coherence faster because of the smaller variation space. Multi-level and top-down selection also perform equally well and reach coherence between 6.000 and 8.000 language games.

Figure 7.18 gives an indication of the kinds of languages that are formed in the population if the agents use the fourth set-up (multi-level selection with lateral inhibition). The graph shows that the generalisation rate of the agents is not really impressive: only three to five generalised roles survive the competition. Moreover, these roles only cover two or maximally three participant roles. This is clear from the fact that there are still 18 to 25 specific markers floating around in the population.

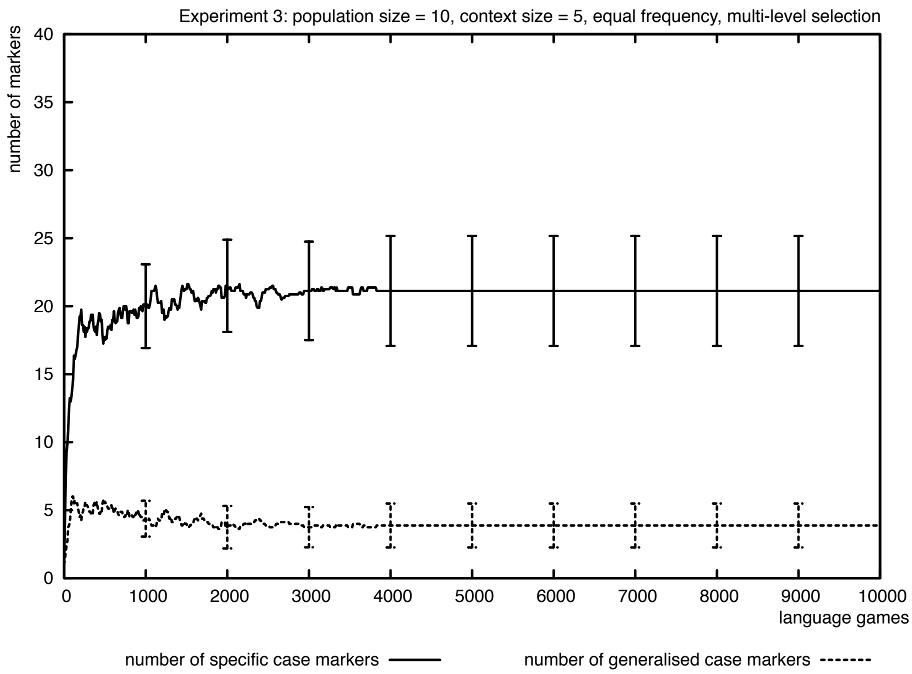


Figure 7.18: This graph shows the average number of markers known by each agent using the strategy of multi-level selection with lateral inhibition (the fourth set-up of experiment 3). The results show that the generalisation rate of the agents is not impressive: only 3-5 semantic roles survive the competition as opposed to 18-25 specific markers.

These results can be compared to the performance of the fifth set-up (multi-level selection with decay) which is illustrated in Figure 7.19. The top graph shows the results for communicative success, cognitive effort, meaning-form coherence and systematicity. As the graph indicates, the agents succeed in reaching full systematicity somewhere between 8.000 and 12.000 language games, which is a bit slower than the alignment strategy using lateral inhibition. The bottom graph shows the average number of markers floating around in the population. Here, we see that the average number of specific markers has made a significant drop from 18-25 markers to only 9. The number of semantic roles shifts from simulation to simulation between 4 and 6. The semantic roles also tend to be more general categories than in the simulations using lateral inhibition.

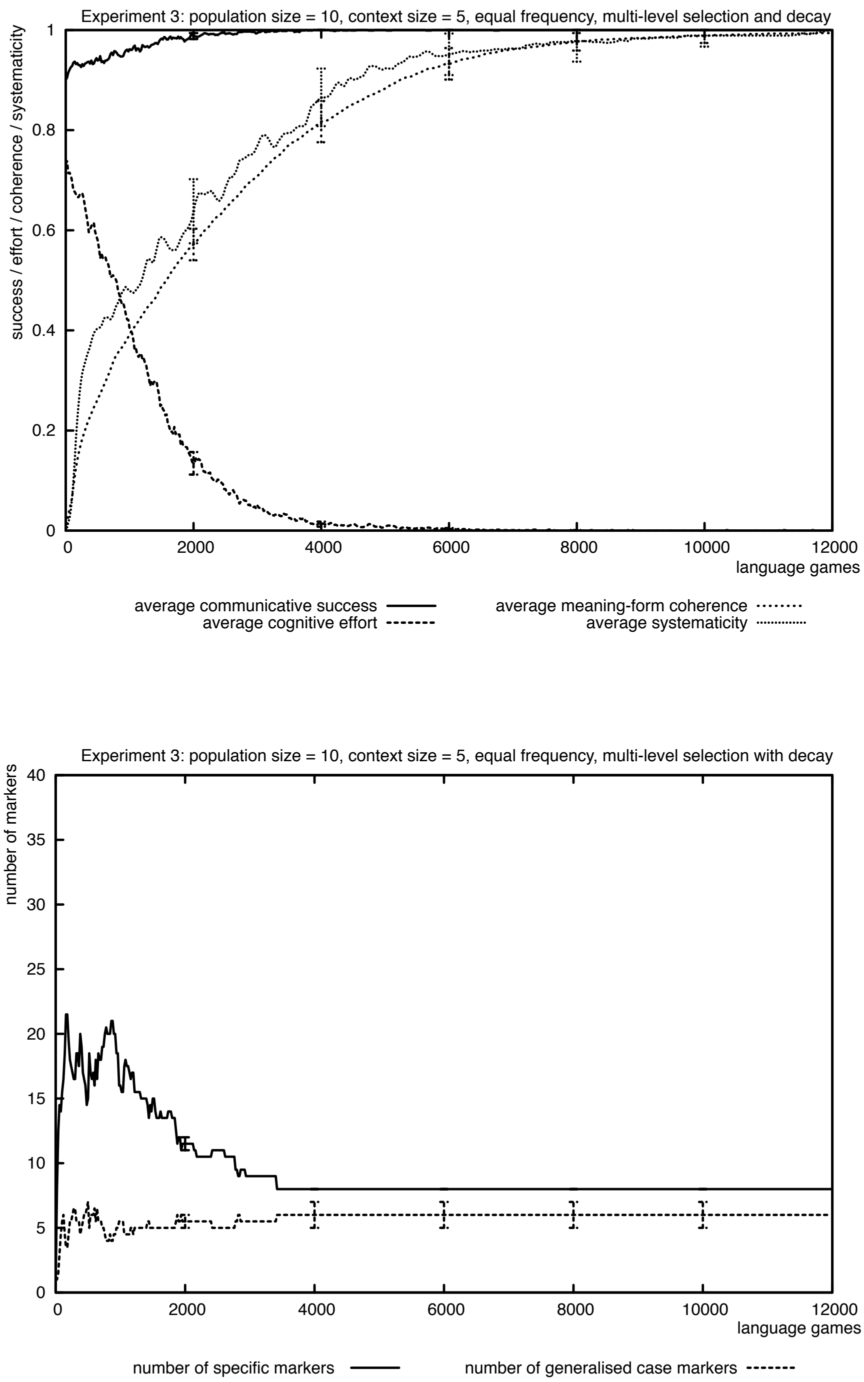


Figure 7.19: The top graph shows communicative success, cognitive effort, systematicity and meaning-form coherence in the set-up using multi-level selection and decay. The bottom graph shows the average number of markers in the same set-up.

Here is a list of markers and the participant roles they cover in one of the simulations (from more general to more specific):

- *-kad:* object-1, approach-1, fall-1, touch-1, move-outside-1
- *-fuir:* grasp-2, hide-2, move-inside-2, touch-2, walk-to-1
- *-kazo:* approach-2, fall-2, grasp-1, hide-1, walk-to-2
- *-hesa:* move-1, take-3
- *-ti:* visible-1, take-1
- *-qiwo:* move-inside-1, give-3
- *-fen:* distance-decreasing-1
- *-rem:* distance-decreasing-2
- *-gaeh:* move-outside-2
- *-wupu:* cause-move-on-1
- *-chuiw:* cause-move-on-2
- *-nuip:* cause-move-on-3
- *-tu:* give-1
- *-hozae:* give-2
- *-fut:* take-2

The above markers occur systematically across patterns for marking the same participant roles. In this specific example, the agents succeeded in reaching coherence and systematicity after only 8.000 language games. When we compare the results to those of baseline experiment 3d, roughly the same level of generalisation is reached.

Finally, Figure 7.20 looks inside the linguistic inventory of a single agent and offers a partial network of the agent's knowledge of its language. The figure concentrates on three constructions (in the middle) which are related to each other (indicated by the dotted line). The relation between the constructions indicate that construction-27 was created as a pattern of construction-2 and construction-10. The figure also shows all the lexical entries that are conventionally associated with these constructions. The links between the lexical entries and the constructions are used for optimising processing: instead of trying out all the constructions in memory, only the linked constructions are considered. Links can be added as part of a problem-solving process during communication or pruned if the co-occurrence is not a successful one. Some redundant co-occurrence links may survive in the inventory. The links can also be seen as fusion links: they are annotated with information on how the participant roles can be fused with the semantic roles of the construction. This annotation is however not used by the agents themselves but for the clarity of interpretation for the experimenter. The actual fusion is taken care of by the unification of the potential valents of the lexical entry with the actual valency of the construction.

Discussion. The results of experiment 3 confirm the problem of systematicity that was uncovered in the other experiments. Here too, the strategy using multi-level selection was the only one to yield fully systematic languages. The systematicity rate was in each of the first four set-ups comparable to the rates in experiment 2. This may come as a surprise since the variation space is potentially smaller because the agents can extend the use of existing markers rather than inventing new ones all the time.

A closer look at the number of markers in the fourth set-up (multi-level selection with lateral inhibition), however, pointed to the reason for the small differences in systematicity rates: only a very small number of semantic roles survived the competition compared to a large number of specific markers. This means that the generalisation rate is not really impressive so the number of variations is not much smaller in these simulations than was the case in experiment 2. The resulting number of specific markers versus semantic roles corresponds to the results obtained in baseline experiment 3c and also the reason for the result is the same: since the competition is held at the level of co-occurrence links and not at the level of constructions, the type frequency of a marker does not translate into a larger category gravity. Competition is held only in a local context so specific markers have an equally high chance of winning as more generalised semantic roles.

The results of the fifth set-up (multi-level selection with memory decay) also confirm the results of the baseline experiments and improve significantly in terms of generalisation over the simulations using the fine-grained lateral inhibition dynamics. The number of specific categories has dropped by half and larger, more general semantic roles have a selectionist advantage because they occur more often. The generalisation rate roughly matches the performance of baseline experiment 3d.

The consistency in number of generalised semantic roles and specific markers across the baseline experiments and the pattern experiments indicate that this is the maximum generalisation rate that the agents can reach. Possible improvements would have to come from two sources:

1. The structure of the world: The capacity of analogical reasoning is heavily dependent on the structure of the world environment in which communication takes place. If the agents have to communicate about lots of events which show recurrent patterns in terms of visual primitives, they will be able to detect more analogies. If, however, the world is totally unstructured, the agents will come up with more specific markers than general semantic roles.

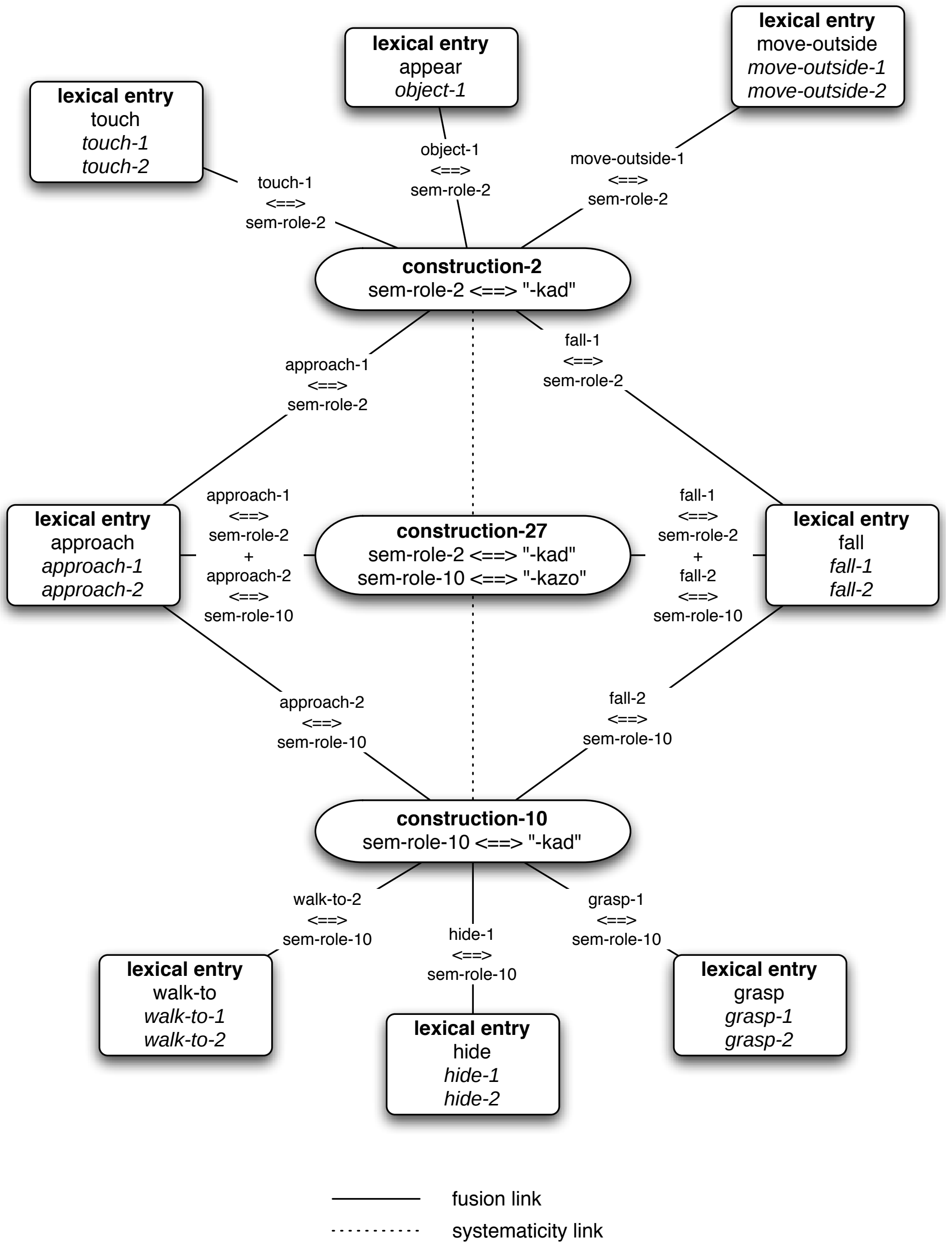


Figure 7.20: This figure gives a partial network for one agent in one of the simulations using multi-level selection with decay. In the middle there are three constructions which are systematically related to each other. The constructions are also linked with lexical entries. The links act both as co-occurrence links for optimising processing and as fusion links for integrating the participant roles of the lexical entries with the semantic roles of the constructions. The networks are constructed in a stepwise fashion as a response to communicative needs.

2. The capacity of analogical reasoning can be made more flexible. At the moment, the agents make a sharp distinction between what is analogous and what is not. A possible relaxation could be to only care about whether the mapping between a source role and a target role is discriminating enough for identifying the target role as well. Another possibility would be to use a similarity or a distance metric instead of the more rigid structural mapping that the agents currently use.

Experiment 3 also shows the potential power of the combination of analogical reasoning, pattern formation and multi-level selection respectively. First of all, analogical reasoning over the linguistic inventory can lead to an increasing generalisation rate in the population, as is also shown in several instance-based approaches to language (Daelemans & Van den Bosch, 2005; Skousen, 1989). These models also argue that a language can look rule-based from outside whereas in fact the generalisation is distributed over the linguistic items in the inventory. This experiment shows that this observation also holds true for the case of the formation of grammar if multi-level selection is applied. Finally, the formation of patterns to improve processing can have dramatic effects on the grammar: the patterns increase the survival chances of its related items; and they may potentially extend their use as well in later interactions.

So far, I did not spend much attention to the efficiency of the formalisation of argument realisation proposed in Chapter 5. In all the experiments presented in this chapter and the previous one, the representation has proven to be flexible enough to deal with the enormous amount of **uncertainty** that is inherent to the formation of new grammar conventions. In this case, the agents needed a flexible way of integrating lexical entries with constructions of various degrees of entrenchment. Instead of copying all the possible case frames into a new entry, the formalisation allowed the agents to constantly 'mould' their lexical entries until a stable set of conventions had been negotiated. In this way, the competition of case markers could be held exclusively at the level of constructions instead of creating an additional competition on the lexical level for how these lexical entries should be integrated with the constructions. The lexical entries also integrated as easily with verb-specific constructions as with verb-class specific constructions.

The formalisation was also flexible enough to deal with **multiple argument realisation**: the agents were capable of integrating a single lexical entry into multiple patterns or constructions without the need for derivational rules or additional copies in the lexicon. Moreover, the lexical entries do not need to 'profile' their participant roles (in Goldberg's sense of profiling, see section 4.1.3 and Goldberg, 1995): the actual valency of a verb is determined by the construction it integrates with. Preferences for certain patterns of argument realisation could be captured in this formalisation by assigning a frequency score to the co-occurrence links of the lexical entries and the constructions that they are conventionally associated with.

One aspect that is still absent in the experiment is how the functions of the case markers start to influence each other once they start combining into patterns. At this moment, the meanings of the markers stay the same and patterning only influences their survival chances. Future work would thus have to include a way for the patterns themselves to evolve, which would also require the analogy to use the patterns as the source domain for innovation rather than focusing exclusively on single markers.

Including the patterns into the search domain could however lead to a huge hypothesis space and a complexity measure is needed to verify whether the algorithm for analogy can scale up to larger worlds while maintaining a reasonable processing time. If not, a possible alternative could be a nearest-neighbour algorithm which has already been successfully applied to various tasks in natural language processing. In a comparison between Royal Skousen's *Analogical Modeling* (Skousen, 1989) and *Memory-Based Language Processing* (Daelemans & Van den Bosch, 2005), Daelemans (2002) shows that a relatively simple and efficient nearest-neighbour learner yields comparable and sometimes even better results than the costly algorithm of Analogical Modeling. This observation seriously challenges the more traditional approach to analogy and is highly relevant for the discussion of this work as well.

7.4.3 Conclusions

In this section I successfully extended baseline experiment 3 to a simulation which includes pattern formation and multi-level selection. The results of experiment 3 confirmed the insights gained from experiments 1 and 2: the problem of systematicity can only be solved if the agents build a structured linguistic inventory and if they exploit this structure through a multi-level selection alignment strategy for converging on a systematic and coherent language. As in the baseline experiments, the best results were obtained using an alignment strategy based on frequency combined with memory decay, which favours more general categories as they have a wider range of usage.

In the discussion of the experiment, I suggested that analogical reasoning (at least during learning and innovation) combined with pattern formation and multi-level selection can lead to a grammar which looks rule-based on the outside, but in which the systematicity actually arises through the distribution of the structured linguistic inventory. A similar argument has previously been made in exemplar-based models of language. I also argued that the formalisation proposed in this thesis is flexible enough to support the experiments and to allow the agents to orchestrate the right kind of selection dynamics.

Finally, I pointed to some shortcomings of the model which need to be addressed in further work. This includes using the patterns as a basis for innovation. I suggested that this could possibly lead to problems in scale-up because of the costly algorithm of analogy and I referred to a nearest-neighbour approach as a possible alternative solution.

7.5 Towards syntactic cases

The experiments so far have dealt with stages 1 to 3 in the development of case markers (see Chapter 4). The next step is the introduction of syntactic roles that group together two or more semantic roles. In this section, I will introduce a first experiment that investigates how the transition to stage 4 can be achieved and what can be learned from the results. I will then use the grammatical square as a roadmap for future experiments.

7.5.1 A first experiment

As I argued in sections 4.2.4 and 4.2.5, syntactic roles impose even more abstraction on the conceptualisation of specific events than semantic roles do. In natural languages, syntactic roles typically emerge when a category gradually starts extending its use until two cases merge into one class. In this first tentative experiment, I will scaffold the merger of cases and assume that a case marker can extend its use by subcategorisation rather than by merging two roles. I will make this assumption more clear in the following paragraphs.

Experimental set-up. The experiment features the exact same set-up as the fourth set-up in experiment 3: the agents are capable of reusing a marker through analogy and combining the markers into larger patterns. They employ the alignment strategy of multi-level selection using lateral inhibition. The main question of the experiment is whether the agents are capable of aligning their grammars, which form an abstract intermediary layer of semantic and syntactic roles which are not directly observable by the other agents.

The novelty of this particular set-up involves the idea of 'reusing as much as possible'. Roughly speaking, a speaker will reuse an existing marker even if it is not analogous to the target role, on the condition that the marker does not cover a conflicting participant role yet. The algorithm is operationalised as follows:

1. If the speaker diagnosed a problem of unexpressed variable equalities, she tries to repair the problem.
2. The speaker checks whether she already has a marker which can be reused:
    (a) Take all known markers. Markers are ordered according to type frequency, that is, from more general (i.e. covering the most participant roles) to less general (i.e. covering the least participant roles). Loop through the markers until a solution is found:
        i. Take all the semantic roles that are covered by the marker, also ordered according to type frequency.
        ii. Loop through the semantic roles until a role is found which is analogous to the target role. If so, return the analogy.
    (b) If a solution is found, return the analogy. If not...

   i. Take the most general marker which does not cover another participant role of the same event yet;
   ii. Create a new role for the target role and make it a subcategory of the chosen marker.
  (c) If there are no markers yet or if no marker can be found that does not already cover a conflicting participant role, create a new marker.

3. The speaker creates the necessary rules and/or links in the inventory.

The hearer learns a marker in a similar way. If she observes a marker in a new situation, she will first try to retrieve an analogous role covered by that marker. If she cannot retrieve the analogy, she creates a new specific role and makes it a subcategory of the marker. It will often happen that the speaker used a marker which according to the hearer already covered a conflicting participant role. For example, the speaker uses *-bo* for marking 'approach-1', whereas the hearer has already observed this marker for covering 'approach-2'. In this case, the hearer will nevertheless create a new specific role for the new use as a possible subcategory of the marker. The fine-grained alignment strategy of the experimental set-up is flexible enough to rule out which participant role should win the competition (unless another marker takes over).

The main idea behind this innovation and learning strategy is that the speaker will be reluctant to invent a new marker. Rather, she will reuse a marker as long as it is discriminating a participant role in an event from the other roles. This means that, in principle, the agents should suffice in using three different markers. This experimental set-up has been implemented in a two-agent simulation and a five-agent simulation.

Results. The two-agent simulation features no variation in the population so the agents have no problems in aligning their grammars since they are endowed with the same algorithm of analogical reasoning. The resulting grammar is illustrated in Figure 7.21 which shows the mapping between participant, semantic and syntactic roles.

The diagram shows that the two agents indeed agree upon three case markers for covering all thirty participant roles. The results even seem to improve on the markers that were formed in baseline experiment 3a: ten semantic roles have been formed as opposed to six specific roles (which are also called 'sem-roles' in the experiment for convenience's sake). On the other hand, the baseline experiment featured two semantic roles which covered six and four participant roles respectively, whereas the semantic roles in this case reach a maximum of four.

In terms of inventory size, the agents require 16 single-argument, 16 two-argument and 3 three-argument constructions. The fact that the number of two- and three-argument constructions is almost the same as in the experiments using no analogy is due to the fact that only in two cases, a larger construction can also be fused with two different

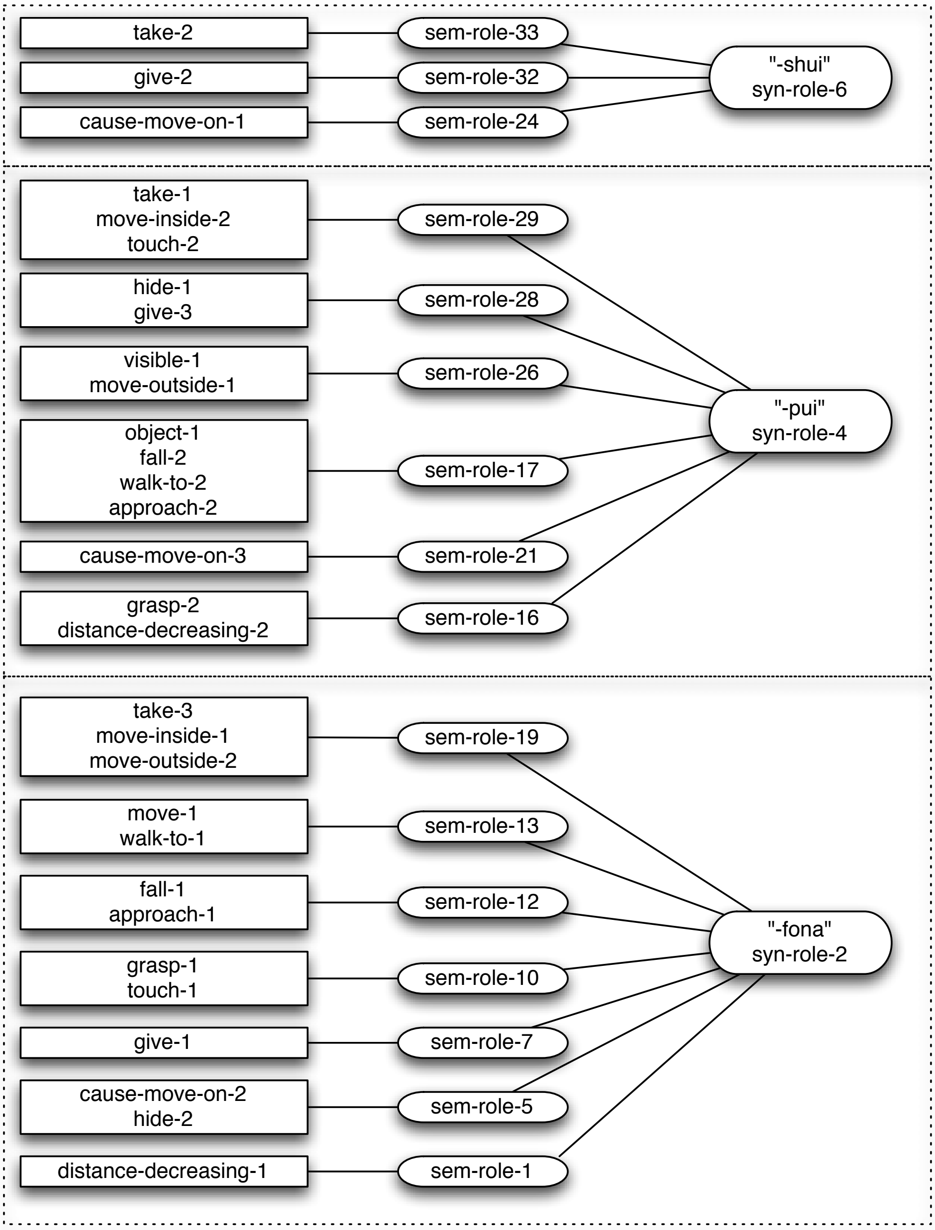


Figure 7.21: This diagram shows the mapping between participant roles and semantic roles; and between semantic roles and syntactic roles in the two-agent simulation on the formation of syntactic roles. Since there are no variations in the simulation, both agents align their grammars perfectly. The results are taken as the baseline for the multi-agent simulations.

lexical entries. For example, *approach* and *fall* integrate with the same three constructions. So even though single-argument constructions can group several participant roles together, the patterns do not succeed in going beyond a verb-specific use.

The results of the two-agent simulation can now be taken as the baseline for the multi-agent simulations. A similar snapshot of the multi-agent simulation is presented in Figure 7.22 and Figure 7.23. These diagrams present the internal mappings between participant, semantic and syntactic roles from two different agents in the same population. Both agents (as well as the other agents in the population) converge on a coherent set of meaning-form mappings.

A closer look at the two diagrams shows that the agents have converged on five different case markers. Two of these markers are participant role-specific, while the others group together multiple roles. Compared to the two-agent simulation, there is a significant loss in number of semantic roles: the first agent has constructed five semantic roles and the second agent has constructed six semantic roles. This leaves 18 and 17 specific roles respectively. Most of the semantic roles also only cover two participant roles.

A comparison of the semantic roles of both agents also shows that they do not align their internal categorisation. For example, the agent in Figure 7.22 groups 'move-outside-2', 'walk-to-1', 'fall-2' and 'grasp-2' together. The other agent has a similar category but has constructed a separate role for 'fall-2'. The other agent has also created a semantic role for covering 'move-inside-2' and 'move', whereas these are two distinct categories for the first agent.

Discussion of the two-agent simulation. The agents in the two-agent simulation were capable of improving over baseline experiment 3a in terms of semantic roles, single-argument constructions and the number of markers. The improvement is due to the fact that the limited use of markers guides the search space more strongly during innovation. In the previous experiments, each semantic role had its own case marker so the chances that they had the same type frequency were quite high. In this case, the speaker would always randomly choose which semantic role to extend. In this new simulation, the type frequency of the syntactic role was more important which led to a faster divergence between the productivity rate of the cases. This means that semantic roles which would otherwise miss extension due to random choice now have more weight to categorise new participant roles.

An interesting side-effect of the innovation algorithm is that there are two syntactic roles which cover almost exclusively semantic roles while a third role acts as a waste basket category for three participant roles which are all three part of events featuring three participants. This maps onto the distinction between agents and patients (and subjects and objects) that is made by most of the languages in the world. Most theories of language assume a (near-)universal distinction between agents and patients

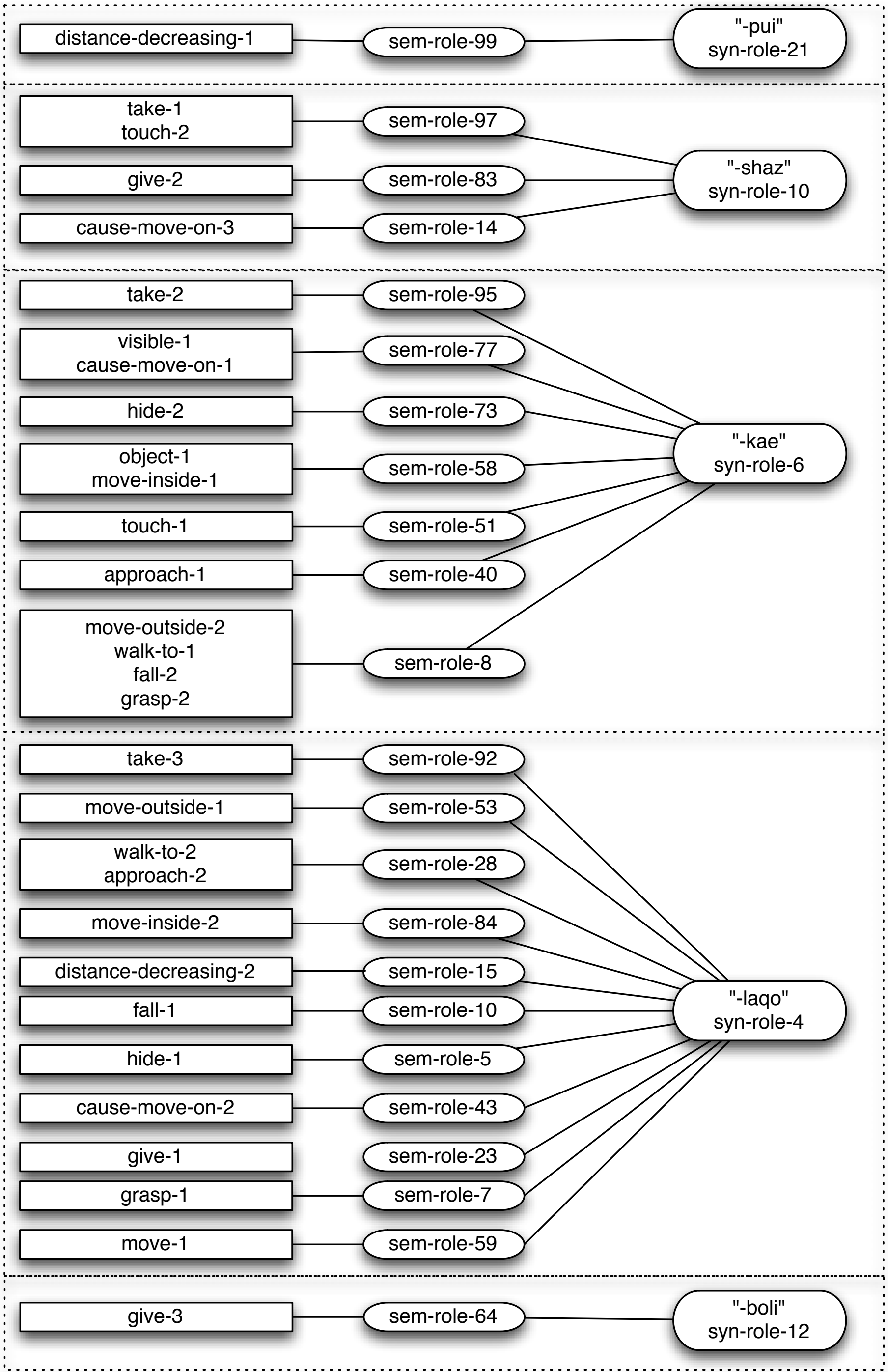


Figure 7.22: The mapping between participant, semantic and syntactic roles in a single agent in the multi-agent simulation.

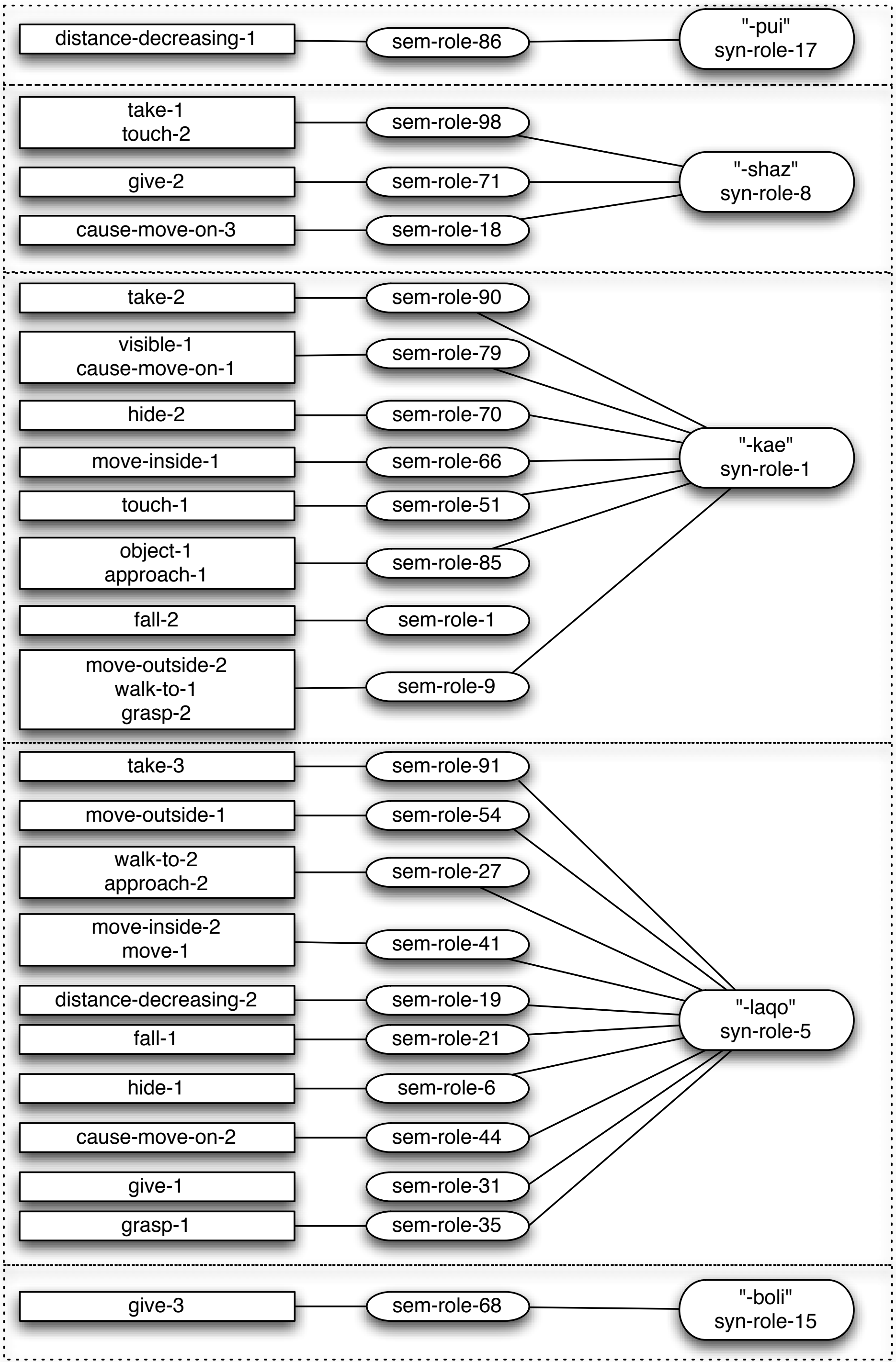


Figure 7.23: This diagram shows the internal mapping as known by another agent in the population.

to be given (either based on a universal conceptual space or on Universal Grammar). This first tentative experiment suggests an alternative hypothesis: the distinction could emerge as a side-effect of communicative goals because language users want to make grammatically and communicatively relevant distinctions. In most of the cases, two or three syntactic roles suffice. The extension of case markers and the merger of semantic role could thus spontaneously lead to 'core' cases.

A final remark considering the two-agent simulation is that there is no gain in terms of inventory size with respect to larger constructions. This is due to the fact that only a couple of constructions share the same semantic roles which combine into the same larger construction. As I suggested during the discussion of experiment 3, the newly formed patterns themselves should be considered by the analogy as well. This could further increase the generalisation and productivity rate of the agents and could be an additional drive towards a prototypical agent-patient distinction.

Discussion of the multi-agent simulation. The results of the multi-agent simulation shows that the alignment of an indirect and multilayered grammatical mapping is no trivial issue: the number of semantic roles constructed by each agent drops significantly and there are differences in how the internal mapping of each agent is organised. This is basically a problem of feedback: there is too much variation floating around in the population for the agents to successfully retrieve the analogy meant by the speaker. Additional feedback could consist of alternative agnating structures (see Chapter 4) which could be exploited for constructing semantic roles. This, however, would require the capacity of dynamically updating the function or meaning of the semantic roles, which the agents do not have (also see the concluding remarks in section 6.4). I will go deeper into this matter in section 7.5.2 in which I discuss the necessary requirements for achieving the transition from two-agent to multi-agent simulations.

Conclusions. In this section, I presented a first experiment towards syntactic roles. In the two-agent simulation, the agents achieved a slightly better categorisation than in baseline experiment 3a. I also suggested that the reuse of case markers to indicate grammatically relevant distinctions could be an alternative hypothesis for the near-universal distinction between prototypical agents and patients. This hypothesis, however, requires additional experiments. In the multi-agent simulation, the agents failed to align their grammars and reach the same generalisation rate as in previous experiments. I argued that this is due to the lack of alternating structures which could offer additional feedback about categorisation on the one hand, and the fact that the agents cannot dynamically update their syntactic cases on the other.

7.5.2 The grammatical square: a roadmap for further work

The first experiments towards syntax showed that once the mapping between semantic roles and syntactic roles becomes indirect and polysemous, the agents are faced with a huge coordination problem: the abstract layer of semantic and syntactic roles is not

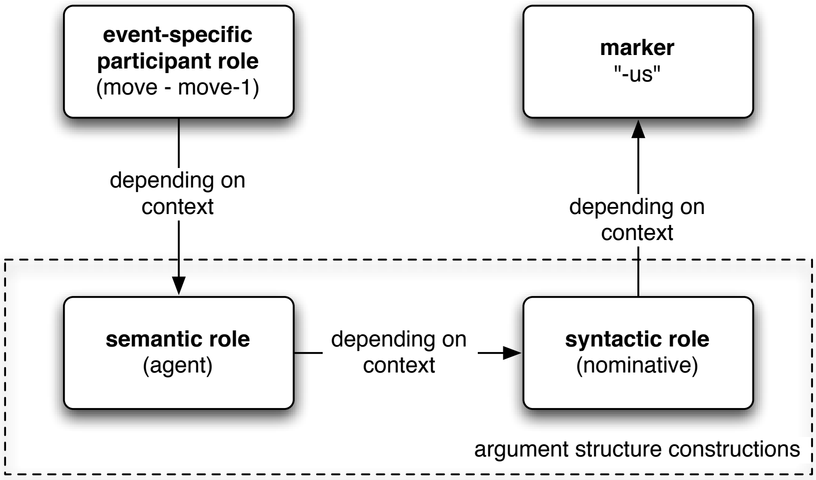


Figure 7.24: The grammatical square can be read as a roadmap for future research. Each mapping in the square is dependent on the context, so experiments should investigate which mechanisms and conditions can lead to such an indirect mapping.

directly observable from the outside and the agents have no means of finding a shared categorisation. Yet the alignment of this internal categorisation is crucial in order to preserve a good productivity and generalisation rate for reaching communicative success in future interactions. In order to take this step, the experimental set-up needs to be seriously expanded. In this section, I will use the grammatical square (repeated in Figure 7.24), as a guidance for identifying which efforts need to be made in order to solve this complex alignment problem.

Mapping participant roles onto semantic roles. Participant roles of a particular event can map onto several semantic roles in natural languages. This is clear in the following sentences, in which *the window* first plays the role of patient, then some entity which undergoes a change of state, and finally some stative entity:

(7.17) He broke the window.

(7.18) The window broke.

(7.19) The window was broken.

In the experiments presented in this thesis, such functional variation was impossible: during conceptualisation, the agents always profile the complete participant role in the event structure. This could lead to sentences similar as example 7.17. However, in the other two sentences, only a subpart of the participant role is profiled: example 7.18 profiles the change of state whereas example 7.19 profiles the resulting state of the participant.

In order to achieve the same functional variation, the agents would thus have to be able to include aspect (and tense) distinctions in their conceptualisation. For this, the conceptualisation algorithm and the algorithm for analogy would have to be changed in order to include the hierarchical structure of the event descriptions and the time stamps provided by the event recognition system. On the level of the interaction pattern, the agents would somehow have to be able to get sufficient feedback in order to recognise and learn the relevant aspectual distinctions. The formation of grammatical markers for aspect (and tense) is no trivial matter and goes well beyond the scope of this thesis.

Mapping semantic roles onto syntactic roles. The mapping between semantic roles and syntactic roles can already be multilayered in nature without taking information structure into account: in examples 7.18 and 7.19, the distinction is not one of active versus passive, but rather one of aspect. In order to see such alternations emerging, the idea of reuse can be exploited again. As the agents develop their grammars, they increase their expressive power. From the moment they want to express aspectual distinctions as well, they could try to reuse the existing grammatical system instead of inventing some new strategy. This puts pressure on the existing conventions and may lead to additional abstractions in the form of syntactic roles.

Another way to investigate the formation of syntactic roles would be to endow the agents with the capacity of dynamically updating the representation of their categories. If categories are not fixed but dynamic, there is a risk of 'category leakage' as observed in natural languages: some categories start expanding their use which could lead to the merger of two semantic roles into one case. Typically, the more frequent cases would start extending their usage which could gradually lead to prototypical agent and patient categories.

Mapping syntactic roles onto case markers. In the experiments, there is a one-to-one relation between syntactic roles and case markers. In natural case grammars, however, there are often paradigms of related markers that together cover a particular case. These marker alternations typically indicate grammatical distinctions in terms of gender and number. Evolving this kind of variation would thus require a need for marking grammatically relevant distinctions between arguments.

Even more intriguing are case systems where the same markers can be used to indicate different cases. For example, Latin uses the inflection *-um* to mark nominative case in neuter singular words (*bellum* 'war') and accusative case in masculine singular words (*dominum* 'master') (Blake, 1994, p. 4–5). This means that case markers somehow manage to grow a paradigmatic case system which goes beyond the borders of individual cases. Applied to the experiments, this would mean lifting the present assumption that competing case markers are always competing with each other for marking a particular participant role. Instead, agents should be allowed to accept competition and variation at each possible mapping in the grammatical square. This may lead to a credit assignment problem in which the agents can never have complete certainty about which mapping in the grammatical square was relevant during a particular innovation.

From the above discussion it should be clear that scaling up the experiments towards richer syntax and grammar is not a trivial matter. It would include research into the emergence of aspect and tense distinctions, a dynamic representation of linguistic categories, and allowing competition on all aspects of the grammatical square. A scale-up to include information structure as well would involve expansion of the language game model to larger dialogues, which creates the need for additional capacities such as episodic memory, scoping and coordination issues and possibly anaphora resolution.

7.6 Summary: a structured inventory

In this chapter I presented a series of experiments which investigated how agents could create larger argument structure constructions by simply combining co-occurring constructions into larger patterns. I argued that natural languages are full of patterns ranging from fully-specified idioms to more schematic constructions. I also argued that pattern formation as a mechanism is relevant to grammaticalization theory and could be a first step towards the grammaticalization of constructions. Next, I offered an operationalisation of pattern formation applied to the case grammar experiment.

In a first experiment, I demonstrated that pattern formation leads to unsystematic languages if the linguistic inventories of the agents are unstructured lists of linguistic items: once a pattern is introduced, it has ‘a life on its own’ and the evolution on the level of patterns may be different than the one on a lower level. This led to languages in which the mapping between meaning and form was inconsistent across patterns.

In a brief review of other related simulations, I indicated that the problem of systematicity has occurred before in other experiments, but that it was never noticed by the experimenter. This led to wrong interpretations of experimental results. One pitfall was making the link between the lack of full systematicity to irregularity in natural languages. I argued that the causes of irregularities in natural languages are fundamentally different and that (given the abstractions found in computational modeling) the agents should reach a stable and fully systematic language.

I then compared various alignment strategies to each other that tried to solve the problem of systematicity. These alignment strategies were inspired by ‘multi-level selection’, a mechanism identified in evolutionary biology which means that being part of a group increases the survival chances of individuals. When comparing the results of the simulations, the only alignment strategy which led to full systematicity across all experimental runs was the multi-level selection strategy in which not only smaller units could benefit from the success of related patterns, but in which also the patterns were rewarded when smaller constructions were successfully applied. Simulations that used only top-down selection displayed a high systematicity rate, but still contained 'frozen accidents'.

The third experiment confirmed the results of experiment 2 and included the capacity of analogical reasoning. The best performance in this experiment was reached using multi-level selection with memory decay in which no explicit lateral inhibition was performed by the agents. I argued that multi-level selection goes against the widespread conceptualisation of the linguistic inventory as a top-down inheritance network and that all levels in the inventory are affected by the evolution of other levels. I also noted that multi-level selection could explain the rise of systematicity without the need for a single abstract rule, but rather from the properties of distributed but related constructions.

The representation that I proposed in Chapter 5 proved to be flexible enough in all the experiments to deal with the uncertainty inherent to the formation of novel grammar conventions and with multiple argument realisation. The main benefit of the representation was that it allowed the agents to orchestrate the competition among variants on the right level and at the same time keep the conventionalised units in the language (in this case the lexical items) adaptive enough to cope with this variation.

Next, I reported a first experiment towards the formation of syntactic cases. The results of the two-agent simulations improved over results of the baseline experiments because the grammar of the agents guided the search space more strongly. I also launched the hypothesis that the near-universal distinction between agents and patients could arise as a side-effect of the growing productivity and generalisation of some categories. In the multi-agent simulations, the agents reached complete communicative succes, but did not manage to align their internal categorisations with each other.

Finally, I discussed some necessary requirements for moving the experiments towards greater complexity so that the agents would be able to construct and share a true case system. I indicated that each mapping in the grammatical square is indirect in natural languages and suggested why this is the case. I argued that additional research is needed which includes distinctions of aspect and tense during event profiling, the dynamic representation of grammatical categories, and many other elements which require a far-reaching rethinking of the experimental set-up. It seems that with the present results, the limits of what can be achieved using the experimental set-up of this thesis have been reached.

Part III

Impact and Conclusions

Chapter 8

Impact on artificial language evolution and linguistic theory

The time has now come to weave through the theoretical foundations of part I and the experimental results of part II to reflect on the contributions of this thesis to artificial language evolution and linguistics. Section 8.1 deals with the first part of this reflection by comparing the results of this thesis to those obtained in a recent study on case marking in the Iterated Learning Model, one of the most widely adopted approaches in the field of artificial language evolution. The comparison shows that the cognitive-functional approach clearly outperforms the Iterated Learning Model and that the work in this thesis is a significant step forwards in the field.

The other three sections of this chapter deal with the contributions of this thesis to linguistics. More specifically, I believe this thesis can have an impact in three domains. First of all, the formalism proposed in Chapter 5 is the first computational implementation of argument structure ever in a construction grammar framework. In section 8.2, I will compare it to an upcoming alternative in Sign-Based Construction Grammar. A second contribution has to do with the structure of the linguistic inventory. The problem of systematicity and multi-level selection is new to linguistics and may have an impact on how we should conceive the constructicon. I will discuss this matter from the viewpoint of construction grammars and usage-based models of language in section 8.3. Finally, the experiments in this thesis and prior work in the field provide alternative evidence in recent debates in linguistic typology and grammaticalization. These debates concern the status of semantic maps and thematic hierarchies as universals of human cognition, and the appropriateness of reanalysis as a mechanism for explaining grammaticalization. Section 8.4 introduces the debates and illustrates how experiments on artificial language evolution can propose a novel way of thinking about the issues at hand.

8.1 Pushing the state-of-the-art

The significance of scientific research can only be fully appreciated by comparing it to other studies in its domain. In this section, I will illustrate how the work in this thesis advances the state-of-the-art in artificial language evolution by comparing it to a recent study by Moy (2006) who investigated the formation of a case grammar in the Iterated Learning Model (ILM), at present one of the most widely adopted models in the field. The comparison reveals some fundamental problems with the ILM and shows that a cognitive-functional approach is the most fruitful way for moving the experiments on artificial language evolution towards greater complexity, expressiveness and realism. In the following subsections, I will summarise four series of experiments reported by Moy and discuss how and why the work in this thesis yields better results. Finally, I will give a general overview of both approaches.

8.1.1 Experiment 1: a primitive case system?

The main objective of Moy's work is to expand the Iterated Learning Model as presented by Kirby (1999a) in order to study the formation of a case grammar. Kirby's original experiments investigated how a recursive word-order syntax can emerge as a side-effect of the cultural transmission of language from one generation to the next without the need for communicative pressures (the so-called 'function independence principle', Brighton *et al.*, 2005a, also see section 3.1.2). Moy's first series of experiments are a replication of these simulations.

The experiment. The experiment features a population of two agents: one 'adult' speaker and one 'child' learner. The adult speaker has to produce a number of utterances that are observed by the learner. The adult speaker has an innovation strategy which allows her to invent a random holistic word for each new meaning if she does not know a word yet. The child learner is equipped with a Universal Grammar in the form of a strong induction algorithm and will try to induce as much grammatical rules as possible. The child will thus overgeneralise the input provided by the speaker which causes language change as illustrated in the child-based model in section 2.3.1. After some time, the adult agent 'dies' and the learner becomes the new adult speaker so its grammar becomes the new convention. A new child learner is then introduced into the population. This population turnover is iterated thousands of times.

The ILM hypothesises that the development of grammar is triggered by the Poverty of the Stimulus or the learning bottleneck: since children cannot observe all possible utterances in a language, there is a pressure on language to become more and more learnable. The linguistic inventory should therefore evolve to an optimal size for a given meaning space. The meaning space consists here of five predicates and five objects which can combine into simple two-argument events such as *loves(john, mary)*. In total there are 100 possible events. The optimal inventory size consists of 11 rewrite-rules: one abstract rule for word order, and ten rules for each word.

As I wrote in section 2.2.1, one challenge for a Universal Grammar mechanism is to provide the agents with a strategy for filtering the 'correct' input from the 'wrong' input. In these experiments, child learners will especially be confronted with conflicting input at the beginning because the language of the adult speaker is still holistic and unstructured. The ILM solves this problem by ignoring all variation. For the learner, this means that once a rule has been induced, all conflicting input is neglected. On top of that, the agents are endowed with a deterministic parser which always picks the first matching rule in the list. Competing rules are therefore never considered because they are lower in the list. This assures a one-to-one mapping between meaning and form which is crucial for the ILM to work properly (Smith, 2003b). The deterministic parser also allows the agents to rely on word order for distinguishing the semantic roles of events.

Results. The results of the simulations show that the agents start inventing holistic utterances which leads to an unstructured language in the first generations. After several hundreds of generations, the language becomes more and more regular and thus learnable due to the overgeneralisation of linguistic input by the child learners. These overgeneralisations become the new grammar of the language once the child replaces the adult speaker. Moy notes, however, that not all languages evolve to the optimal size of 11 rules, but that *"a significant number of the runs converged on a larger grammar with 16 rules"* (p. 113). These grammars contain two distinct noun categories: one for the agent and one for the patient. Here is an example of such a grammar which uses an SOV word order:

	s/ [P, X, Y]	→	3/X, i, 1/Y, 2/P
	1/anna	→	i, p, l
	1/kath	→	c, s
	1/mary	→	t, a
	1/john	→	j, e
	1/pete	→	h
	3/kath	→	a, k, f
	3/pete	→	a, u, f
(8.1)	3/mary	→	t, s
	3/anna	→	g
	3/john	→	p
	2/kisses	→	t
	2/hates	→	z, s
	2/loves	→	m, q, j
	2/adores	→	u, i
	2/sees	→	m, y
	(Moy, 2006, p. 113)		

In the above grammar, the agents have to use a different word for the same object depending on which semantic role it plays in the event:

(8.2)	*p*	*i*	*ta*	*mqj*
	john	[EMPTY]	mary	loves
	'John loves Mary.'			

(8.3)	*ts*	*i*	*je*	*mqj*
	mary	[EMPTY]	john	loves
	'Mary loves John.'			

This kind of 'suboptimal' grammar is not exclusive to Moy's replication experiment: it also occurs in Kirby's original simulations. For example, Kirby (1999b) suggests that the two distinct noun categories can be considered as case-marked nominals. In the discussion of her replicating experiment, Moy seems to follow this hypothesis:

> Could we view such a grammar as exhibiting some form of primitive case system, in that it is possible to distinguish subject forms of nouns from objects, rather than using the same form for both? This is analogous perhaps to highly irregular forms of case found in some languages, such as the English pronouns *I, me, we* and *us*, where the nominative forms, (*I* and *we*) used to the [sic] represent the subject of a sentence, have no morphological relationship to the accusative forms used for objects (*me* and *us*). (Moy, 2006, p. 114)

Discussion. As I argued in section 7.2.3, considering the above grammar as some kind of primitive case system is an over-interpretation of experimental results. While there are many attested examples in natural languages in which a word form depends on the linguistic context (e.g. many Slavic languages such as Russian use different lexical entries for verbs depending on aspect), they do not come about as frozen accidents of the learning mechanism. As for the English pronouns, they are remnants of a stage in the development of English where the grammar had a fully productive case system. Both Moy and Kirby thus fail to identify the problem of systematicity that occurs in the experiments.

This observation illustrates the importance of a strong dialogue with linguistics. Most research in the field so far has contented itself with shallow comparisons to natural languages based on the more philosophical sides of the debates I introduced in Chapter 2. It is however crucial to go into more empirical details in order to appreciate the enormous complexity involved in grammatical phenomena in natural languages. In this thesis, I tried to offer such an appreciation of case markers in Chapter 4. A better understanding of the developmental pathways of case markers helped to uncover the problem of systematicity in this thesis and prevented me from simply concluding that the lack of systematicity could be mapped onto some of the more exotic case alignment systems found in the world's languages (see Figure 8.3). Other first steps towards domain-specific dialogues exist for colour terms (Steels & Belpaeme, 2005), spatial language (Loetzsch *et al.*, 2008) and vowel systems (de Boer, 1999).

In this thesis, I solved the problem of frozen accidents through multi-level selection. I implemented multi-level selection as an alignment strategy which is therefore tightly coupled to communicative success. In the ILM, however, communicative success has no impact on the behaviour of the agents so multi-level selection cannot be readily applied here. Second, multi-level selection only makes sense if there is variation in the model, but the ILM avoids variation as much as possible so once a frozen accident occurs, it is hard to get rid of it.

Even though it is unwarranted to equate the frozen accidents in the ILM experiments with case systems in natural languages, Moy's work takes an interesting turn by asking how the ILM can be expanded so that it favours these 'suboptimal' languages. If Moy succeeds in demonstrating the factors that *systematically* lead to the formation of such languages, her experiments could still come up with relevant pressures for evolving a case language. I will describe her attempts in the following sections.

8.1.2 Experiment 2: dealing with variation

In order to encourage the emergence of a primitive case system, Moy (2006, chapter 5) experiments with various modifications of Kirby's parser that make it less deterministic and which allows for variation in word order. The hypothesis is that if the agents can no longer rely on word order for distinguishing agents from patients in the events, they will start learning grammatical rules with case-like properties. I will again summarise the obtained results and discuss them afterwards.

Experimental results. Moy first tries to introduce variation in the word order by allowing the agents to randomly choose among conflicting rules or by reshuffling the inventory. In all the simulations, however, the agents fail to converge on a compositional grammar: the linguistic inventories of the agents are very large and too much conflicts are known for reaching a regular language. Moy argues that this is due to the fact that compositionality can only emerge in the ILM if variation in meaning-form mappings is ignored: learners will not consider any new variation anymore once they have associated a certain meaning with a certain form, and the deterministic parser excludes the use of variations. The same conclusion has also been suggested by Smith (2003b), but as opposed to Moy who rejects this unrealistic assumption, Smith argues that natural languages have such a bias towards one-to-one mappings as well.

Moy thus points to a fundamental problem with the ILM: in order to allow variation in word order, the agents need to be capable of parsing and producing competing rules. Yet, in order for the grammar to emerge at all, the ILM expects that a single variation is maintained. Moy notes that in order to dampen the search space, the agents need a way to prefer some variations over other ones. She therefore endows the agents with an alignment strategy which takes the frequency of rules into account. The more frequent the rule, the higher the probability that the agent will select it for production. The alignment strategy allows the agents to reach a compositional language again. It does not, however, lead to an increase in the number of primitive case grammars. In additional experimental set-ups, Moy allows even more word order freedom by reshuffling sentences before they are actually produced, but this again does not lead to a preference for the primitive case grammar. Finally, manipulating the size of the transmission bottleneck (i.e. the number of utterances that the child learner observes during a lifetime) does not yield significant results either.

Moy concludes that the small effect of free word order and the bottleneck size is due to the fact that the agents do not have to reach mutual understanding: if the free word order cannot be parsed by the hearer, it will simply be ignored so it will not lead to a change in the linguistic inventory. In other words, the child agent does not really care about converging on the same language as that of the adult speaker but learns whatever hypotheses its induction algorithm comes up with. Moy argues that the agents thus have no need for disambiguating semantic roles which prevents them from reaching a case-like grammar.

Discussion. Moy's experiments reveal a number of fundamental shortcomings of the ILM: first of all, the agents have no way of dealing with variation. This problem did not surface in Kirby's prior work because he implemented a bias towards one-to-one mappings that excludes the possibility of competitors. Moreover, the ILM is typically implemented using only two agents, so no competing rules can ever be introduced in the population. Moy's results clearly show that there is no connection whatsoever between the grammar that is acquired by the learner and the grammar of the speaker: in fact, the learner will apply a 'first come, first serve' approach to learning grammar in which the first successful parse leads to a fixed entry in the inventory. This means that the agents in Kirby's models did not *learn* to mark the distinction between semantic roles, but that they have this distinction already built in.

Moy rightfully notices that the agents need some kind of (alignment) strategy in order to reach a regular language. She introduces an utterance-based strategy which counts the number of occurrences of rules. Equipped with this alignment strategy, the agents are capable of producing and parsing multiple word orders, but this has hardly any effect on the language of the agents. Moy then rightfully concludes that the agents need **communicative pressures**: since the child learner does not care about communicative success, any grammar induction will do. The experiments thus show that the 'function independence principle' cannot lead to grammar (at least not a case grammar) unless by accident or through a Universal Grammar. The transmission bottleneck is therefore not a sufficient trigger for marking event structure through grammar.

Problems with multi-agent simulations. Another way in which the problems of the ILM can be demonstrated is to scale up the experiments to multi-agent populations. This has indeed been attempted by Smith & Hurford (2003). Smith & Hurford reached the following conclusions:

1. The agents fail to align their grammars because they have no alignment strategy and different agents will come up with different innovations and generalisations;

2. As a result, learners are presented with inconsistent training data;

3. The eager abstraction algorithm has disastrous consequences for the performance of the agents.

Smith & Hurford consider the option of allowing the learners to keep multiple hypotheses. Since the agents are however 'eager learners', their abstractions harm their performance: abstractions made by one agent are not always the same as the abstractions made by another one so the agents never reach a shared language. This means that the agents would have to maintain all possible grammars, which rapidly becomes intractable. Another problem is that the agents need to find out which grammar is 'best'. Smith & Hurford refute the possibility of a cost system (similar to the lateral inhibition strategies proposed in many problem-solving models) on the grounds that it is 'rather ad hoc'. The solution offered by Smith & Hurford is however at least ad hoc as most cost systems are: they implement strong production biases coupled to 'smart pruning' in order to reduce the number of hypotheses.

As I argued in Chapter 2, however, variation is a fundamental property of language. Introducing additional production biases is a way to put less weight on the cultural evolution of language and more on the genetic endowment of the agents, which is exactly the contrary of what the ILMs try to show. Moreover, there now exist mathematical models of lateral inhibition dynamics which solve the 'rather ad hoc' status of cost systems (Baronchelli *et al.*, 2006; De Vylder, 2007). Also the alignment strategy based on token frequency proposed in this thesis is rooted in proposals made in cognitive-functional models of language.

The real problem for the ILM is however that such a cost system only works in a bottom-up, instance per instance learning of the grammar. Otherwise it would indeed lead to an intractable search space containing all possible grammars. In order to avoid innate biases towards one-to-one mappings, it is therefore necessary to introduce **an utterance-based selectionist system** rather than a grammar-based selectionist system. The work in this thesis demonstrates how such a bottom-up approach can lead to systematic languages if analogy is used in combination with multi-level selection.

8.1.3 Experiment 3: implementing communicative pressures

Following the conclusion that the agents need additional communicative pressures in order to form a primitive case grammar, Moy (2006, chapter 6) presents a series of experiments in which ambiguous word orderings occur. The hypothesis is that this ambiguity will lead the agents to prefer rules that use different noun categories to mark the distinction between semantic roles. I will review the most important experiments here.

Experimental results. In a first set-up, Moy implements an 'inversion procedure' that swaps the ordering of words for which the agent already has a rule. This procedure thus guarantees ambiguity during learning. The results are however disastrous: most of the runs fail to reach any kind of regularity at all. Moy assigns the reason for this failure to the fact that the child learner indeed faces ambiguity, but that she still does not need to reach mutual understanding with the speaker. The hearer thus keeps ignoring the ambiguous utterances if they do not fit the grammar rules that have already been acquired.

In an attempt to fix this problem, Moy implements a learner that does not tolerate ambiguity and an interaction script that forces the speaker to introduce unambiguous utterances: the meaning that was parsed by the hearer is compared to the intended meaning of the speaker. If there is a mismatch, the speaker has to come up with an alternative verbalisation. However, this implementation does not lead to success either: in most cases, the speaker does not have an alternative way to verbalise an utterance so she will invent a new holistic string. Since the ILM features meaning transfer, the hearer will each time learn this new utterance. The result is that there is constant innovation in the simulations so the agents never reach an 'optimal' language. Additional attempts, such as punishing ambiguous utterances, do not yield improvement either.

Discussion. Even though Moy's experiments started from a correct observation, the 'communicative pressures' that are needed for a case grammar were not operationalised and implemented in a satisfying way. The main problem is that the agents in her experiments are not truly communicating. First of all, the speaker's output was randomly changed by an artificial procedure in order to create ambiguities for the hearer. This is already highly problematic because it would require some malicious mind-reading from the speaker's part, but there are other more serious issues: the speaker does not have any communicative goal at all. She just produces an utterance but does not care about whether this utterance had the intended effect in the hearer's mind. Other simulations that emphasize the importance of communication have all come to the conclusion that the speaker must try to produce an utterance in such a way as to improve the chance of being understood by the hearer (Smith, 2003a; Steels, 2003b). In this thesis, this was operationalised in the form of re-entrance (see Chapter 6).

From the part of the hearer, there is the same problem. 'Ambiguity' in Moy's model does not mean that the hearer did not understand what the speaker said because there is meaning transfer and the hearer can even compare her parsed meaning to the speaker's intended meaning. The problem is that the hearer does not attempt to align her grammar to that of the speaker at all: a mismatch in meanings means the rejection of the speaker's utterance. This means that the hearer stubbornly sticks to her induced grammar rules which may well be completely different than the grammar of the adult speaker because of the greedy induction algorithm. This problem does not occur in the experiments in this thesis because the agents try to find out what the speaker's intentions were and want to conform to the conventions of the population.

In short, none of the agents ever actually try to reach communicative success. As I argued in Chapter 6, language is an **inferential coding system** in which the language users are assumed to be intelligent enough to make innovations that can be understood by the hearers, and in which hearers can make **abductions** about the speaker's intended meaning. In Moy's experiments, the agents only want to get rid of internal inconsistencies rather than trying to converge on the same grammar.

8.1.4 Experiment 4: more innate knowledge

In a final series of experiments, Moy (2006, chapter 7) tries to address the problem of the 'suboptimal' primitive case system in a different way. She starts from the observation that the grammar inducer *"is not capable of* effectively *learning inflectional grammars"* (p. 206). For example, the default inducer would fail to notice the inflectional markers in an utterance such as *johnalovesmaryb*: instead of recognising *-a* as an indicator of the subject and *-b* as an indicator of the object, the learner will induce them as an integral part of the word form. Moy therefore proposes several implementations that attempt to solve this problem. I will discuss one of these attempts in which she implements a richer meaning representation.

Experimental results. In all of the previous experiments, the semantic roles of agent and patient were implicit in the meaning so the inducer wasn't able to extract them. Moy therefore decides to make the meaning representation richer by explicitly adding the roles of 'actor' and 'actedon'. The meaning [loves, john, mary] would thus become [[act, loves], [actor, john], [actedon mary]] (p. 217). The 'optimal' case marking language should have 15 rules: one top-level rule, ten lexical entries, two markers and two rules to combine the markers with the noun categories. Despite the explicit representation of semantic roles, however, not all simulations led to satisfying results. There were two kinds of problematic cases:

- One type of grammar again features two completely distinct noun categories with different words for the same meaning. The inducer failed to recognise inflectional markers for the agent and the patient.
- A second type of languages *did* have inflectional markers, but they also featured two different lexical entries for the nouns. An example of such a grammar is:

	s/ [P, X, Y]	→	1/Y, 13/X, 6/P
	13/ [A, B]	→	16/B, 15/A
	1/ [C, D]	→	3/C, 8/D
	15/ actor	→	i
	3/ actedon	→	h
	16/ john	→	v, s
	16/ jane	→	k, h
(8.4)	8/ john	→	e, v, s, n
	8/jane	→	s
	6/ [E, F]	→	14/F, 7/E
	7/ act	→	i, b
	14/ adores	→	y, n, k
	14/loves	→	j, v
	14/sees	→	h, m, l
	(Moy, 2006, p. 230)		

This would lead to sentences such as:

(8.5) *h* *evsn* *kh* *i* *jv* *ib*
actedon john jane actor loves act
'Jane loves John.'

(8.6) *h* *s* *vs* *i* *jv* *ib*
actedon jane john actor loves act
'John loves Jane.'

Discussion. Moy's experiments suggest that the Iterated Learning Model cannot overcome its bias towards one-to-one mappings which is confirmed in the fact that experimenters working with the model increasingly turn to innate constraints for explaining grammar formation: Moy explicitly includes two semantic roles in the meaning space, and Smith (2003b) and Smith & Hurford (2003) implement explicit biases towards one-to-one mappings. This is highly problematic since grammatical categories are clearly multifunctional. It also means that the ILM studies have given up on their original objectives: instead of demonstrating that grammar can evolve through *cultural* selection, a strong Language Acquisition Device is built in.

The experiments in this thesis do not presuppose such a bias towards one-to-one mappings and in fact offer **the first multi-agent simulations ever featuring polysemous categories**. By taking communicative pressures seriously and by providing the agents with a richer cognitive apparatus including analogy and multi-level selection, this thesis demonstrated how agents can deal with the variety and uncertainty that is inherent to multi-agent simulations, how they can self-organise and coordinate a grammar involving multifunctional semantic roles, and how they can reuse the same linguistic items in multiple patterns of argument realisation.

What is also striking about Moy's results is the fact that even though there are only two agents at each given time and even though the learner is equipped with a highly specialised learning mechanism, the grammars can still get stuck in a state which is not fully systematic. This suggests that (a) a learning bottleneck is not a sufficient pressure for reaching a systematic language, and that (b) an innate mechanism that seriously restricts the space of possible grammars is not sufficient for dealing with variation. The only way to avoid this unsystematic state, as I argued in Chapter 7, is to assume a cognitive-functional view on grammar in which agents are given credit for possessing the right skills and alignment strategies to arrive at a shared and systematic communication system.

8.1.5 Summary: case markers serve communication

In this section I reviewed experimental results on the formation of a case grammar obtained by Moy (2006). I showed that in its present state, the Iterated Learning Model fails at developing a case grammar and that the reasons for this failure provide additional evidence that the cognitive-functional approach is the best way of doing research on the formation of grammar. The comparison also shows that the work in this thesis offers a significant step forward in the field of artificial language evolution.

The ILM and the problem-solving approach both argue that grammar evolves through cultural evolution but both models diverge significantly in terms of assumptions and hypotheses (see section 3.1 and Table 8.1). The most important difference is that the ILM tries to explain as much grammatical structure as possible as a side-effect of the cultural transmission of language from one generation to the next, whereas the problem-solving approach assumes that grammatical development is triggered by the need to optimise communicative success. This leads to two different types of learners: in the ILM the learner needs strong innate constraints as opposed the problem-solving approach where the learner is equipped with a rich cognitive apparatus for detecting and solving communicative problems.

Since the learner in the ILM is endowed with some kind of Universal Grammar, the model has a lot of difficulties with Tomasello's 'linking problem' (see section 2.2.1): inconsistent input data and variation in the population can only be handled by strong innate biases towards one-to-one mappings and by innate linguistic categories. The problem-solving model of this thesis, on the other hand, has no problems with variation and inconsistent input data. It features a redundant and bottom-up approach to language and an utterance-based selectionist system in which careful abstraction is possible despite the enormous uncertainty about the conventions in the speech community. The agents in this thesis do not assume one-to-one mappings and the experiments offer the first multi-agent simulations ever which involve polysemous categories.

Another achievement of this thesis is that it detected and solved the problem of systematicity. I showed that the problem also occurs in Iterated Learning Models (and other models, see Chapter 7) but that it remained unnoticed so far. I argued that this was due to an underestimation of the complexity of natural language phenomena and a shallow comparison between natural and artificial languages. The work in this thesis is more firmly rooted in linguistic theory and offers a model which is closer to attested cases of grammaticalization.

It should be noted that both approaches are not mutually exclusive. The problem-solving approach naturally incorporates all the learning constraints that are focused upon in the ILM but goes much further in terms of learning and innovation strategies in order to optimise communicative success. Expanding the model to multi-generational population dynamics is indeed technically quite trivial and has already been successfully demonstrated in many experiments (e.g. De Beule, 2008; Steels & McIntyre, 1999). Expanding the ILM to a multi-agent population, however, is much more problematic as was demonstrated by Smith & Hurford (2003). The work in this thesis therefore seriously challenges the widely adopted ILM and shows which direction has to be taken to increase the complexity and expressiveness of the artificial languages.

	Moy (2006)	**This thesis**
Triggers of grammar	- Poverty of the Stimulus - Learning bottleneck - Function independence	- Communicative success - Reducing cognitive effort - Increasing expressiveness
type of learner	- Eager learner - Greedy induction - Top-down	- Lazy learner - Careful abstraction - Bottom-up
Language change	- Hearer-based innovation - Mismatch in learning	- Speaker-based innovation - Adoption by hearer - Propagation in population
Grammaticalization	- Analysis only - Holistic strategy - Start without language	- Analogy and patterns - Continuity principle (reuse) - Start from lexicon
Inventory	- Rewrite rules - Ordered but unstructured	- Constructions - Structured through usage
Meaning space	- Two-way contrast - Semantic roles given	- Event-specific meaning - No prior semantic roles
Population	- Two-agent simulation - Generational turnover - Speakers vs learners	- Multi-agent simulation - Single generation - Peer-to-peer interactions

Table 8.1: This table summarises the main differences between Moy (2006) and the work in this thesis.

8.2 Argument structure and construction grammar

From this section onwards I will discuss the contributions of this thesis to linguistics, starting with the issue of representing construction grammar. In Chapter 5 and van Trijp (2008a), I proposed a formalisation of argument realisation in Fluid Construction Grammar. Even though my proposal was explicitly implemented for supporting experiments on artificial language evolution, some of its ideas can be relevant for formalisms of natural languages as well. Within the family of construction grammars, however, no alternative computational implementation has been reported yet in a peer-reviewed publication that can handle both parsing and production. As such, van Trijp (2008a) offers the first computational implementation of argument structure constructions within a construction grammar framework.

This doesn't mean that no other work has been done yet: at the moment this thesis was written, a different proposal was still in the process of being implemented in Sign-Based Construction Grammar (SBCG, Fillmore *et al.*, in preparation), a formalism that combines HPSG (Pollard & Sag, 1994) with Berkeley Construction Grammar (BCG, Kay & Fillmore, 1999). Since the draft proposals are closely related to the more construction-oriented approaches in HPSG, I will assume here that SBCG can be implemented and applied in a *computational* formalism as well. In the remainders of this section, I will briefly illustrate how SBCG deals with argument structure (based on Fillmore *et al.*, in preparation; Kay, 2005; Michaelis, to appear; Sag, unpublished). I will then illustrate this for the English ditransitive based on Kay (2005) and then compare it to my own representation of argument structure.

8.2.1 Argument structure in BCG and SBCG

Early representations of argument structure in a construction grammar framework (such as Goldberg, 1995) and the representation used in this thesis propose argument structure constructions in which the meaning can be seen as a skeletal or schematic event type such as X-CAUSES-Y-TO-MOVE. These constructions then have to unify or fuse with the lexical entry of the verb. In later versions of BCG and in SBCG, however, argument structure constructions are implemented as two-level derivational rules that have a 'mother'-component (MTR) and a 'daughter'-component (DTR). The DTR unifies with the lexical entry of the verb and is complemented by the MTR. The SBCG proposal looks very much like lexical rules in lexicalist accounts (see section 4.1), but it does not entail the conservation of thematic structure (Michaelis, to appear).

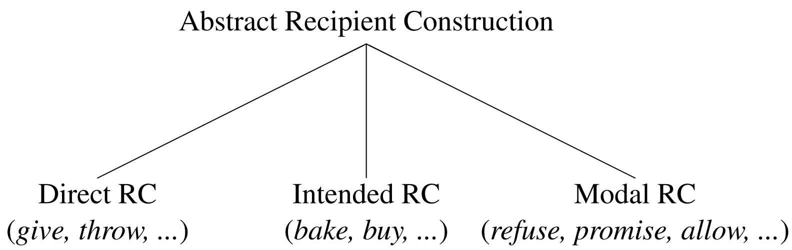
The SBCG approach is in the same spirit as Goldberg (1995) in the sense that minimal lexical entries are assumed for verbs which have to be complemented by argument structure constructions. SBCG however goes a step further and allows argument structure constructions to *override* the default behaviour of a verb (Michaelis, to appear; Sag, unpublished). Sag (unpublished) writes that SBCG has a feature called 'argument structure' (ARG-ST) which encodes the valence of a verb. This feature is a structured list which is coupled to the 'Accessibility Hierarchy' of Keenan & Comrie (1977): the first argument maps onto subject, the second onto the direct object, etc. The rank-based listing of arguments is chosen to eliminate the need for explicit features such as 'subject' and 'object'.

Different argument realisation patterns, such as the active-passive alternation, are represented through different values for the ARG-ST list. SBCG has two different ways to implement these differences: either a derivational construction overrides the default ARG-ST of the verb (as is the case in passivisation) or there is lexical underspecification (for example in locative alternations). In addition to the feature ARG-ST, SBCG also has a feature VAL(ence) which is a list of the syntactic elements that a linguistic expression yet has to combine with. Sag gives the example of the verb phrase *persuaded me to go*, which takes the VAL list < NP > because it still needs to combine with a subject NP. The clause *my dad persuaded me to go* takes an empty VAL list because it doesn't need to combine with any other argument anymore. Only lexical constructions have both ARG-ST and VAL features. Phrases, clauses, NPs, and other items only have empty VAL lists.

8.2.2 An example: the ditransitive construction

I will now illustrate argument structure constructions in SBCG based on Kay (2005)'s analysis of the English ditransitive. Kay's proposal is somewhat different than the most recent developments of the SBCG architecture, but the underlying ideas are the same (Kay, pers. comm.). It shares many aspects with the approach offered by Goldberg (1995), such as the assumption that there is a default and minimal lexical entry for each verb. For example, the lexical entry of the verb *to bake* contains two minimally required arguments (a baker and a thing that is baked). Argument structure constructions can add arguments such as the beneficiary in *she baked him a cake*. As explained in Chapter 4, Goldberg sees argument structure constructions as larger patterns carrying grammatical meanings which have to be 'fused' with the meaning of the lexical entries. Kay, however, proposes argument structure constructions which are more like lexical constructions with a 'mother constituent' and a single daughter. The daughter unifies with a lexical entry and is elaborated by the mother constituent. Applied to the English ditransitive construction, Kay proposes three 'maximal recipient constructions' which all three inherit from a more schematic 'Abstract Recipient Construction':

(8.7)



These constructions are represented in a unification-based grammar as detailed in Kay & Fillmore (1999), and which is similar to analyses of argument structure in HPSG (Pollard & Sag, 1994). Here is the Abstract Recipient Construction:

(8.8)

$$\begin{bmatrix} \text{syn} & \begin{bmatrix}\text{cat v, lex +, min-}\end{bmatrix} \\ \text{sem}\,|\,\text{cont} & \boxed{0}\begin{bmatrix} \text{handle} & \boxed{1} \\ \text{index} & \boxed{4} \\ \text{list} & \left\langle \begin{bmatrix} \textit{intnl-act} & \\ \text{handle} & \boxed{1} \\ \text{actor} & \boxed{2} \\ \text{undergoer} & \boxed{3} \\ \text{intnd-rslt} & \textit{handle} \\ \text{event} & \boxed{4} \end{bmatrix}, \begin{bmatrix} \textit{receive} & \\ \text{handle} & \textit{handle} \\ \text{theme} & \boxed{3} \\ \text{recipient} & \boxed{5} \\ \text{event} & \textit{handle} \end{bmatrix}, ... \right\rangle \end{bmatrix} \\ \text{valence} & \left\langle \begin{bmatrix} \text{syn} & \text{NP} \\ \text{instance} & \boxed{2} \end{bmatrix}, \begin{bmatrix} \text{syn} & \text{NP} \\ \text{gf} & \text{obl} \\ \text{instance} & \boxed{3} \end{bmatrix}, \begin{bmatrix} \text{syn} & \text{NP} \\ \text{instance} & \boxed{5} \end{bmatrix} \right\rangle \end{bmatrix}$$

$$|$$

$$\begin{bmatrix} \text{syn} & \begin{bmatrix}\text{cat v, lex +}\end{bmatrix} \\ \text{sem}\,|\,\text{cont} & \boxed{0} \\ \text{valence} & \left\langle \begin{bmatrix} \text{syn} & \text{NP} \\ \text{instance} & \boxed{2} \end{bmatrix}, ... \right\rangle \end{bmatrix}$$

The Abstract Recipient Construction shows all information which is common to all English ditransitive constructions. The lower box represents the DTR constituent which needs to unify with the lexical entry of the verb. As can be seen, its valence list contains an NP (which plays the role of 'actor') and an under-specified argument (...). The top box represents the MTR constituent which complements the lexical entry of the verb with the 'recipient' role. The 'list'-feature displays the construction's primary semantic frame (*an intentional act* which has an actor and an undergoer) and additionally an *intended result* which still needs to be specified by another frame. Kay proposes that the intended result in the Direct RC is the 'receive frame' (see below) whereas in the other sub-constructions it is not.

The Direct RC corresponds to Goldberg (1995)'s 'central sense' of the ditransitive as in *she gave him a book* (also see example 4.25). The main frame of the MTR component is a CAUSE-MOVE act which is defined as a subtype of an intentional act. The receive frame is unified with the intended result of the main frame indicating that there was an actual transfer of possession. The daughter's valence list indicates that it takes verbs which have at least two arguments (an actor and an undergoer). The Direct RC can be illustrated as follows:

(8.9)

$$\left[\begin{array}{ll}
\text{syn} & \left[\text{cat v, lex +, min-}\right] \\
\text{sem}\,|\,\text{cont} & \boxed{0}\left[\begin{array}{ll}
\text{handle} & \boxed{1} \\
\text{index} & \boxed{4} \\
\text{list} & \left\langle \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{cause-to-move}} \\
\text{handle} & \boxed{1} \\
\text{actor} & \boxed{2} \\
\text{undergoer} & \boxed{3} \\
\text{intnd-rslt} & \boxed{6} \\
\text{event} & \boxed{4}
\end{array}\right], \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{receive}} \\
\text{handle} & \boxed{6} \\
\text{theme} & \boxed{3} \\
\text{recipient} & \boxed{5} \\
\text{event} & \boxed{4}
\end{array}\right] \right\rangle
\end{array}\right] \\
\text{valence} & \left\langle \left[\begin{array}{ll}
\text{syn} & \text{NP} \\
\text{instance} & \boxed{2}
\end{array}\right], \left[\begin{array}{ll}
\text{syn} & \text{NP} \\
\text{gf} & \text{obl} \\
\text{instance} & \boxed{3}
\end{array}\right], \left[\begin{array}{ll}
\text{syn} & \text{NP} \\
\text{instance} & \boxed{5}
\end{array}\right] \right\rangle
\end{array}\right]$$

$$|$$

$$\left[\begin{array}{ll}
\text{syn} & \left[\text{cat v, lex +}\right] \\
\text{sem}\,|\,\text{cont} & \boxed{0} \\
\text{valence} & \left\langle \left[\begin{array}{ll}
\text{syn} & \text{NP} \\
\text{instance} & \boxed{2}
\end{array}\right], \left[\begin{array}{ll}
\text{syn} & \text{NP} \\
\text{instance} & \boxed{3}
\end{array}\right], \ldots \right\rangle
\end{array}\right]$$

The Intended RC corresponds to Goldberg's sense 'F' (see example 4.30) as in *she baked him a cake*. In this sense of the ditransitive, there is no actual transfer of possession entailed but only the intention of transfer. A second reason for positing a different construction for the Intended RC is that it cannot occur in the passive. Kay therefore includes an explicit stipulation in the construction which states that it cannot combine with a derivational passive construction. All the remaining senses of the ditransitive are grouped together in the Modal RC. This construction is similar to the Intended RC in that it does not entail actual transfer either, but it is different in the sense that it doesn't require *beneficiary* semantics. Kay argues that specific meanings are contributed by the verbs themselves so no additional constructions need to be posited.

8.2.3 Discussion and comparison

In this section I will compare the argument structure approach in Sign-Based Construction grammar to my own proposal in Fluid Construction Grammar. I will indicate what the problematic issues are with SBCG and how the FCG solution could have an impact on future construction grammar representations.

Derivational versus non-derivational constructions. The most remarkable distinction between argument realisation in Sign-Based Construction Grammar and Fluid Construction Grammar is that FCG sticks with the traditional construction-based approach in which argument structure constructions have skeletal meanings that need to unify or fuse with the semantics of the lexical entries of the verbs. FCG could thus be said to implement a 'fusion' process similar to proposals made by Goldberg (1995). SBCG, on the other hand, has given up on this kind of analysis and moved closer towards HPSG by using (almost lexical) derivational rules which feature two components: a mother and a daughter.

These different approaches are the result of different solutions to the same problem: multiple argument realisation (see section 4.1). Both SBCG and FCG try to solve the problem through the notion of 'potential syntactico-semantic arguments' (Sag, unpublished) or what I called 'potential valents' in Chapter 5. This word 'potential' refers to the fact that a lexical entry can combine with multiple argument realisation patterns. However, the implementation of this 'potential' is fundamentally different in both formalisms.

SBCG starts from the traditional assumption that lexical entries have a fixed predicate-frame which is implemented in the ARG-ST list. In order to still cope with multiple argument realisation patterns, this ARG-ST list either needs to be under-specified or overridden by derivational rules. This capacity of overwriting the predicate-frame of a verb is in fact the only difference with traditional lexicalist accounts. FCG, on the other hand, does not assume a minimal lexical entry: a linguistic item merely lists its potential from which more grammatical constructions can select the actual valency. In other words, the meaning of the verb strongly influences its morpho-syntactic realisation but its actual valency is still dependent on the other constructions that combine with it.

The difference can be easily explained through an analogy to mathematics which I borrowed and adapted from Michaelis (to appear). Michaelis writes that constructions have the possibility to change the associations within an arithmetic sequence like *2 x (3 + 4)* to the sequence *(2 x 3) + 4*, which would yield different results (*14* and *10*). The individual numbers, however, denote the same value in each sequence. Michaelis' analogy does not quite fit, however, since in SBCG a number would be listed with a minimal ARG-ST (for example saying that '2' has to be used in a sum). SBCG therefore does more than storing the entry '2' with its denotation and would need a derivational rule which overrides the specification of '2'. FCG, on the other hand, would list the number '2' and state that it can be potentially used in sums, divisions and other functions without actually committing to a single operation. The construction would then pick what it needs from the number.

Evidence from corpus-linguistics. Technically speaking, the implementation differences between SBCG and FCG do not really matter. The main problem with derivational constructions, however, is the assumption that some senses of lexical items are more central or more basic than others. Traditionally, these 'minimal entries' are however based on intuition rather than empirical data. For example, what is the minimal entry for a verb such as *to give* if the following examples are taken into account:

(8.10) He gave her the book.

(8.11) She was given the book.

(8.12) He gave blood.

(8.13) Give it!

(8.14) Give it to me!

(8.15) Give me the book!

(8.16) Women always give and give, while men only take.

More examples can easily be found. The point is however that FCG would have no real preference for either pattern (except perhaps as the result of frequency and priming effects) since the lexical entry does not contain a fixed predicate-frame. SBCG would list *give* as a three-place predicate even though numerous sentences can be observed in which not all three arguments are present. Both formalisms thus make different predictions as to the frequency of argument realisation patterns: FCG allows for different frequency patterns for each verb individually (which is captured through co-occurrence links between lexical entries and constructions), whereas SBCG predicts a most basic use of a verb with derived and therefore less frequent uses.

These predictions can be verified through careful corpus studies. A good example is the active-passive alternation. In FCG, the passive is an argument structure construction in its own right which stands on equal footing with active argument structure constructions. In SBCG, on the other hand, the passive construction is a derivational construction which needs to overwrite the default active ARG-ST.

The relationship between active and passive has been investigated by Stefanowitsch & Gries (2003) in a 'collostructional analysis'. Collostructional analysis combines statistical data of co-occurrences between words with a close attention to the constructions in which these words occur. This method allows for a detailed analysis of the relations between words and constructions. Stefanowitsch & Gries use a slightly extended collostructional analysis for investigating various alternations among which the active-passive one. The results of their study shows that there are clear semantically motivated classes of distinctive collexemes for both the active and the passive. The most distinctive collexemes with respect to the active voice were *to have* along with emotional-mental stative verbs such as *think, say, want* and *mean*. With respect to the passive voice, there is a clear class of verbs that *"overwhelmingly encode processes that cause the patient to come to be in a relatively permanent end state"* (p. 110), such as *base, concern* and *use*.

Stefanowitsch & Gries thus confirm prior claims by Pinker (1989) and conclude that the passive construction is primarily a semantic construction rather than a construction that is mainly used for marking differences in information structure. In other words, there are no empirical grounds for assuming that the active construction is basic and that the passive construction has to be derived from it. This observation is highly problematic for the lexical derivations in SBCG but can be captured nicely by FCG.

Thematic hierarchy. The above observations also make the ranked ordering of the ARG-ST of SBCG highly problematic: many verbs apparently prefer the passive construction and therefore violate the ranking more frequently than they follow it. Moreover, the universality of the thematic hierarchy (or the 'Accessibility Hierarchy' which SBCG also inherited from HPSG) has become a matter of big debate due to the serious empirical problems with such hierarchies (see the discussion in section 4.1.2 and Levin & Rappaport Hovav, 2005, chapter 6). In fact, many researchers in HPSG have turned to macro-roles or other constructs to get rid of the unsatisfactory thematic hierarchies (see section 4.1.2 and Davis & Koenig, 2000).

As I already suggested in Chapter 4, FCG does not assume any notion of thematic hierarchies, macro-roles or universal linking rules. Instead, preference patterns for argument linking and realisation emerge as a side-effect of analogy in innovation and multi-level selection: analogical reasoning explains why existing linguistic items get reused in new situations and multi-level selection assures a growing systematicity in which linguistic items combine into a structured linguistic inventory. Recurrent patterns are hypothesised to be captured by the distributed constructions which show systematicity in their classification behaviour. I will come back to this matter in more detail in the other sections of this chapter.

The formation of argument structure constructions. A third problem with the architecture of SBCG is that it would be very hard to implement in the scenario of an emergent language or in which the first formation of grammar takes place. If there is no grammar yet, there are simply no conventions to build a grammar upon. Language users would nevertheless have to agree on the 'basic' argument structure of a lexical entry despite the many conflicting variations floating around in the population (which are often more frequent than what would be intuitively speaking the 'minimal' entry). Then they would have to agree somehow that all possible alternations are in fact derivational constructions. On top of that, these derivational constructions have to be nicely ordered: for example, the passive alternation comes after the ditransitive derivation which comes after the lexical entry. Again, FCG seems to be more flexible in this respect. Even though the experiments in Part II did not reach the full complexity of natural languages, the formalism has proven itself to be very effective even though there were already quite complex internal structures and multiple levels of linguistic organisation involved.

8.2.4 Summary: contributions of this thesis

In this section, I discussed the Sign-Based Construction Grammar approach to argument structure constructions and compared it to argument realisation in Fluid Construction Grammar. I first showed that SBCG has been strongly influenced by HPSG: the traditional construction-based approach has been exchanged for a two-level analysis with derivational constructions which have the capacity of overwriting the predicate-frame of a lexical entry. I also briefly illustrated SBCL for the English ditransitive. In the comparison between SBCG and FCG, I argued that FCG provides novel insights which are not captured by SBCG and which could be adopted in future implementations.

First of all, FCG does not need derivational rules but allows each linguistic item to follow an individual pathway of development and does not require a lexical item to take a fixed predicate-frame. I pointed to a corpus study on the active-passive alternation in English to argue against the intuitive and pre-theoretical notion of a 'minimal' lexical entry that is proposed in SBCG and most approaches in linguistics. SBCG therefore only technically improves on lexicalists accounts by allowing predicate-frames to be overridden (and thus allowing some patterns of argument realisation which are impossible in those lexicalist accounts), but in practice it keeps the same flaws.

I also argued that the thematic hierarchy of SBCG cannot be maintained for the empirical reasons discussed in Chapter 4 and because of the implausible architecture of minimal lexical entries with fixed predicate-frames. I further suggested that this architecture would be very hard to implement in a scenario in which the grammar still has to be formed as was the case in the experiments in this thesis. The FCG representation, on the other hand, dealt quite effectively with the enormous uncertainty of grammatical conventions that occurred in the simulations.

The representation offered in this thesis can thus have an impact on linguistic theory in the sense that the assumptions about minimal lexical entries should be given up in favour of a more flexible approach. I have demonstrated that this can be achieved within the experiments using careful abstractions: linguistic items which have a more 'free' distribution (typically lexical items) can list their potential valents and create co-occurrence links with other items. The more constrained items (typically grammatical constructions) then select from this potential what they need. Another possible alternative could be exemplar-based models in which no abstractions are made but where the instances are kept for processing. Both exemplar-based models and the approach in this thesis feature systematicity through the distribution of linguistic items which may seem rule-gouverned on the outside. Finally, I suggested that the systematicity in my model comes about as a result of analogical reasoning combined with multi-level selection. Analogy can explain systematicity in innovation behaviour, whereas multi-level selection further streamlines the language as an alignment strategy.

8.3 Analogy, multi-level selection and the constructicon

In generative grammar, the problem of systematicity has never been an issue since all categories and grammar rules are assumed to be innate. In construction grammars and usage-based models of language, however, the linguistic inventory is supposed to be acquired in a bottom-up fashion so more attention has been given to how this inventory should be structured. In this section, I will give a brief overview of the most important proposals in construction grammar based on Croft & Cruse (2004, p. 262–290). Next, I will argue that construction grammars need to take multi-level selection into account in how they conceive the relations between constructions in the linguistic inventory or the 'constructicon'.

8.3.1 The organisation of the linguistic inventory

Croft & Cruse (2004) write that the structured, linguistic inventory of a speaker *"is usually represented by construction grammars in terms of a* **taxonomic network** *of constructions. Each construction constitutes a* **node** *in the taxonomic network of constructions*" (p. 262). In other words, constructions are related to each other through taxonomy links or instance links which describe a relationship of schematicity between two constructions. The following example shows a taxonomic relation between the idiom *The X-er, the Y-er* and an *instance* of that more schematic construction:

(8.17) [The X-er, the Y-er]
|
[The bigger they come, the harder they fall.]

(Croft & Cruse, 2004, p. 263, example 3)

The rule of thumb for deciding when a construction has its independent node in the network is when not all aspects of the construction's semantics or syntax can be derived from its subparts or from more schematic constructions. For example, the idiom *to kick the bucket* has its own representation in the network because its meaning cannot be derived from the individual words in combination with a schematic transitive construction. Most but not all theories assume that *to kick the bucket* is also part of the inheritance network but which locally overrides the default behaviour of the more schematic constructions:

(8.18) [VerbPhrase] (Croft & Cruse, 2004, p. 263, example 4)
|
[Verb Obj]
|
[[*kick* Obj]
|
[*kick* [*the bucket*]]

Depending on how much redundancy the theory allows, frequent instances can be kept as well even though a more schematic construction may already exist (also see section 2.4.3). Finally, sentences usually feature *multiple inheritance*: constructions often only offer a 'partial specification' of the grammatical structures of their daughter constructions. Croft & Cruse give the example *I didn't sleep*, which inherits from both the [Subject - Intransitive Verb] construction and the [Subject Auxiliary-n't Verb] construction (p. 264, example 6).

The most influential construction grammars all assume the above organisation of the linguistic inventory. Croft & Cruse discuss four of them: Berkeley Construction Grammar (Kay & Fillmore, 1999), the Lakoff/Goldberg model (Goldberg, 1995), Cognitive Grammar (Langacker, 1987) and Radical Construction Grammar (Croft, 2001). The latter three are also considered to be usage-based models of language. Croft & Cruse compare the different models based on a couple of questions, of which the following two are directly relevant for our discussion (p. 265, questions (iii) and (iv)):

1. What sorts of relations are found between constructions?
2. How is grammatical information stored in the construction taxonomy?

Berkeley Construction Grammar. In the discussion of Berkeley Construction Grammar (BCG) in section 8.2.1 I already briefly mentioned that BCG features an inheritance network for organising the linguistic inventory. Unlike examples 8.17and 8.18, however, BCG does not allow for any kind of redundancy. It is a complete inheritance model in which information is only represented once and at the highest, most schematic level possible. This also means that BCG does not require all constructions to be symbolic units (i.e. form-meaning mappings): they can be entirely syntactic or semantic as well.

BCG therefore captures all information in terms of taxonomy links. Since no information is stored more than once, parts of constructions can in fact be children of other parent constructions. The network therefore not only features instance links between constructions, but also between parent constructions and parts of other constructions.

The Lakoff/Goldberg model. The model proposed by Lakoff (1987) and Goldberg (1995) focuses more on the categorisation relations that may exist between constructions. Next to the taxonomy/instance links, Goldberg also proposes a meronomic or subpart link (p. 78) and a 'polysemy' link (p. 38). The subpart link is different from the BCG subpart links: in BCG, a subpart is a complete instance of a more schematic construction, whereas Goldberg sees subpart links as constructions which are subparts of larger constructions but nevertheless have an independent representation in the inventory. The 'polysemy' links are links between constructions that have the same syntactic specification but different semantics. I already illustrated this polysemy link for the various senses of the English ditransitive construction in section 4.1.3.

One important aspect of the Lakoff/Goldberg model is the notion of a prototype and (metaphorical) extension. For example, if constructions are related through polysemy links, there is always a 'central sense' assumed. For the English ditransitive, this is the sense of actual transfer as in *I gave him a book*. Goldberg and Lakoff propose a somewhat different model when it comes to metaphorical extension: Goldberg assumes that metaphorical extension involves a superordinate schema from which the central sense and the extenstion(s) are instances; Lakoff does not assume such a schema.

The type of inheritance in the Lakoff/Goldberg model is different from BCG in the sense that an instance is allowed to locally overwrite some information that is normally inherited from a higher schema. For example, a schematic category such as BIRD may contain the feature FLIES, but this is not true for penguins. In the penguin-category, the inheritance of FLIES is therefore blocked by local specifications. This solution is also handy when there is conflicting information in the case of multiple inheritance: the instance is then assumed to be represented as a full entry in the inventory.

Cognitive Grammar. Langacker's Cognitive Grammar (CG) is regarded by most cognitive linguists as some form of construction grammar because it shares many of its assumptions and objectives. Langacker assumes that a category typically has a prototypical member or a set of members and that new instances are categorised by extension from the prototypes. Next to this model of prototypes and extension, CG also allows for a more schematic unit which subsumes the prototype and its extensions. This view comes closest to the model of extension through analogy that I operationalised in Chapter 6.

The organisation of the linguistic inventory is dependent on language use. The entrenchment or independent representation of a linguistic item is hypothesised to depend on its token frequency: if a unit occurs frequently enough, it is stored in memory. Productivity of a linguistic unit goes hand in hand with its extension through language use: if a (prototypical) category gets extended to new situations, it increases its type frequency and hence its productivity. As said before, categories can form a network based on prototypical members (instances) and non-prototypical members, but there are also abstractions which are related to their members through taxonomy links.

Radical Construction Grammar. The word 'radical' in Radical Construction Grammar (RCG) comes from the fact that RCG does not assume constructions to be built from atomic categories such as nouns or verbs, but rather that the construction is the atomic unit of language. All other categories are defined in terms of the constructions they occur in. Categories are thus assumed to be construction- and language-specific. For example, the transitive construction and intransitive construction are hypothesised to contain two different verb categories: the transitive verb and the intransitive verb. The superordinate category Verb is seen as a linguistic abstraction over those two categories (Croft & Cruse, 2004, p. 287–288):

(8.19) [MVerb-TA]

[IntrSbj IntrV] [TrSbj TrV TrObj]

In the above example, the label MVerb is used to indicate that this is a morphological construction (TA stands for Tense and Aspect). The superordinate abstraction can only be made if it is linguistically motivated. For example, both transitive and intransitive verbs can be marked for tense and aspect so both categories should be able to occur in those Tense-Aspect constructions. In short, RCG is a strongly non-reductionist approach as opposed to BCG.

RCG features the same taxonomy links as other construction grammars and also allows for redundant information according to the principles of usage-based models of language. One other important aspect of RCG is that it is based on the 'semantic map' model (see section 8.4). In this model, all constructions are hypothesised to map onto contiguous regions in 'conceptual space' which is assumed to be universal. Finally, syntactic structures are defined as language-specific units but in relation to 'syntactic space' which aims at typologically comparing the world's languages.

8.3.2 The linguistic inventory in Fluid Construction Grammar

Before I start the comparison between Fluid Construction Grammar and the above theories, I would like to emphasize again that FCG takes a design stance towards the formation of grammar and that it therefore only implements mechanisms that are experimentally demonstrated to be necessary requirements. The fact that FCG does not make the same abstractions or does not feature the same complex mechanisms for organising the linguistic network therefore does not mean that they are refuted, but only that they are not necessary (yet) for the level of complexity that is reached in current simulations. On the other hand, FCG can show which proposals stand the computationally rigid test in less complex languages. Secondly, by demonstrating novel but necessary mechanisms in those less complex languages, FCG can show which ideas are currently being overlooked by linguistic theories.

The formation of linguistic categories. With respect to the 'atomic' building blocks of a grammar in formation, FCG is closest related to Radical Construction Grammar. From an evolutionary point-of-view, it is more natural to think of constructions or form-meaning mappings as the atomic units in language and that other categories are dependent on the organisation of these constructions. For example, the experiments in Part II do not feature an explicit category for nouns or verbs, yet all words can be used without any problem in argument structure constructions. Further categorisations should be functionally motivated. For example, if the agents should also worry about tense and aspect marking, they might need additional generalisations over their existing constructions.

This scenario is attractive in many ways. First, the agents do not need to agree on a set of building blocks such as nouns or verbs before they can start combining them into sentences. Instead they keep on constructing new categories on the fly but only when this optimises communication and thus when it is functionally motivated. This approach also seems to fit natural languages better since it is impossible to come up with an abstract rule that can be applied to all parts of speech of a language. Finally, this approach also suits my proposal of potential valents for linguistic items: the freer and typically lexical items can be potentially used in many different constructions, whereas the more grammaticalized, tightened constructions typically decide on the actual valency of a linguistic expression.

FCG therefore rejects the reductionist approach of BCG (and SBCG). Reductionist approaches are still dominant in linguistics as a result of a desire for maximising 'storage parsimony' in the linguistic inventory. Croft & Cruse (2004, p. 278), however, point to psychological evidence that suggests that storage parsimony is a cognitively implausible criterion for modeling the linguistic inventory. Language users rather seem to store a lot of redundant information which requires more memory but which optimises 'computing parsimony' because not all information has to be computed online.

Innovation through analogy and pattern formation. Fluid Construction Grammar also subscribes to the usage-based model and argues that innovation occurs through analogical reasoning. In the experiments in Part II of this thesis, I implemented an innovation strategy in which the productivity of a category is related to its type frequency and which is therefore similar to proposals made in Cognitive Grammar. FCG also allows for careful abstraction in which an instance link is created between the more abstract category and the specific instances that are compatible with it. A second drive for innovation is pattern formation: frequently co-occurring utterances are stored as independent units in memory. This is also completely in line with usage-based models that take token frequency as an indicator of entrenchment. The newly formed patterns themselves may be extended through analogy as well.

One salient feature of FCG is that all innovation occurs in a stepwise fashion. If a careful abstraction is made, it is at that moment only valid for the instances that were used in creating the abstraction. The newly formed category therefore does not automatically extend its use to other situations: an explicit link in the network has to be created during other interactions. For pattern formation as well, links are kept between the newly created pattern and its subparts. All the links in the inventory are used for optimising linguistic processing: instead of considering the entire memory, only linked constructions are unified and merged. Only when this strategy leads to communicative problems, the language user will try to adapt the inventory through analogy.

Multi-level selection in the emergence of language systematicity. The work in this thesis has also uncovered the problem of systematicity which has so far been overlooked by all linguistic theories. The usage-based models presented in the previous section mainly focus on a top-down inheritance network and seem to assume that this suffices for reaching and maintaining systematicity if the network is combined with an innovation strategy based on type frequency and productivity.

The experiments in Chapter 7, however, demonstrate that this is not the case. Next to an innovation strategy which systematically reuses productive and successful items of the inventory, language users need an alignment strategy based on multi-level selection to further streamline their inventories and keep the generalisation rate of their language high. The experiments demonstrated that a top-down strategy does not suffice but that the success and evolution of specific instances must also have a way to influence the more schematic constructions in the network. The networks therefore need **systematicity links** rather than (only) taxonomy links.

On the status of inheritance networks. The experiments on multi-level selection show how a linguistic network similar to the one proposed in Radical Construction Grammar could gradually emerge: a nonreductionist approach is taken in which each construction has its specific categories. However, FCG does not make explicit generalisations over these constructions as is done in RCG, but rather keeps systematicity links which are used by the multi-level selection alignment strategy. There is no need for an inheritance network and all utterances are licensed by unifying and merging fully specified constructions.

This architecture suffices for the kinds of experiments performed in this thesis and only further work can show whether additional abstractions and perhaps inheritance networks are really needed. A serious challenge to these kinds of abstractions and inheritance networks is posed by the successful application of instance-based models in natural language processing such as Memory-Based Language Processing (Daelemans & Van den Bosch, 2005) and Analogical Modeling (Skousen, 1989). Another challenge for inheritance networks, I believe, is that they might require abstractions that are too greedy and therefore harmful to the communicative success of language users, especially in experiments on the formation of grammar. As becomes very clear in such experiments, agents have to deal with an enormous amount of uncertainty about the conventions in their population. It might very well turn out to be that a fully redundant model (with or without careful abstraction) using multi-level selection is a more adequate model.

8.3.3 Summary: contributions of this thesis

In this section I summarised the different views on the structure of the linguistic inventory of some current theories and compared them to the approach adopted in Fluid Construction Grammar (also see Table 8.2). Even though FCG is applied to the formation of artificial languages, the formalism and the experiments can be seen as a testbed for verbal theories of natural languages as well.

The comparison made clear that the FCG approach shows strong affinity for redundant and usage-based theories such as the Lakoff/Goldberg model, Cognitive Grammar and Radical Construction Grammar. One contribution made by this thesis is therefore one of the first implementations of such a usage-based model in the field of artificial language evolution. Within the setting of the experiments, I have shown that many of the proposed mechanisms can be successfully operationalised such as analogy, pattern formation, and type and token frequency.

Next to confirming the potential of these mechanisms in a computational model, the experiments also indicated that systematicity in a language can only emerge if the language users have an alignment strategy based on multi-level selection. Multi-level selection favours constructions that are systematically related to each other and make sure that local developments can have an impact on the entire linguistic inventory. I argued that verbal theories in linguistics have overlooked this mechanism so far and have focused too much on top-down processing involving some model of inheritance. I therefore argued for systematicity links instead of or along with taxonomy links.

The implementation I presented in this thesis did not involve inheritance networks that are proposed by most construction grammar theories. The main reason is that the experiments did not require them so they are not built in. As I explained at the beginning of this section, FCG tries to limit its cognitive apparatus to only those mechanisms that are demonstrated to be necessary through computational simulations and robotic experiments. Further experiments are therefore needed to see whether inheritance networks are indeed essential for structuring the linguistic inventory. I pointed to a possible problem with inheritance networks in the extreme situation of the formation of grammar: they might require too greedy abstractions which may be harmful for the linguistic performance of language users.

	Approach	**Constructional relations**	**Network organisation**
BCG/SBCG	Reductionist	- Taxonomy links - Meronomic links	Complete inheritance
Lakoff / Goldberg	Usage-based	- Taxonomy links - Subpart links - polysemy links	Default inheritance
CG	Usage-based	- Taxonomy links - Prototypes	Inheritance as categorisation
RCG	Usage-based	- Taxonomy links - Meronomic links - Prototypes	Semantic map model
FCG	Usage-based	- systematicity links - Co-occurrence links	- No inheritance - Multi-level selection

Table 8.2: This table summarises the different approaches to the organisation of the linguistic inventory in some current theories of construction grammar.

8.4 Linguistic typology and grammaticalization

The previous two sections mainly dealt with the relations between construction grammar and the experiments in this thesis. In this section, I will discuss how the methodology of artificial language evolution can provide novel insights to the fields of grammaticalization and linguistic typology. The experiments in this thesis are not conclusive enough yet to yield a strong and operational hypothesis, but at least they offer an alternative explanation in some of the debates that I will explore here. More specifically, I will show how computational modeling can provide additional evidence in the debates on the status of semantic maps and thematic hierarchies. Finally, I will also comment on future applications for grammaticalization theory.

8.4.1 The status of semantic maps

Semantic maps have offered linguists an appealing and empirically rooted methodology for visualising the multifunctional nature of grammatical categories and for describing recurrent structural patterns in how these functions relate to each other. Consider the following examples in which various functions of the English preposition *to* are illustrated along with some corresponding examples from the French preposition *à* and the German dative case (taken from Haspelmath, 2003, example 2, p. 212 and example sentences on p. 213–215):

(8.20) English preposition *to*:

a. Goethe went to Leipzig as a student. (direction)

b.	Eve gave the apple to Adam.	(recipient)
c.	This seems outrageous to me.	(experiencer)
d.	I left the party early to get home in time.	(purpose)
e.	This dog is (mine/*to me).	(predicative possessor)
d.	I'll buy a bike (for/*to) you.	(beneficiary)
e.	That's too warm (for/*to) me.	(dative judicantis)

(8.21) French preposition *à*:

Ce	*chien*	*est*	*à*	*moi.*
this	dog	is.3SG	to	me

'This dog is mine.' (predicative possessor)

(8.22) German dative case:

Es	*ist*	*mir*	*zu*	*warm.*
it	is.3SG	I.DAT	too	warm

'It's too warm for me.' (dative judicantis)

The universality of semantic maps. Instead of listing the various functions of a grammatical morpheme or 'gram', semantic maps offer a *"geometrical representation of functions in "conceptual/semantic" space that are linked by connecting lines and thus constitute a network" (Haspelmath, 2003, p. 213).* Figure 8.1 gives an example of a semantic map which shows some typcal functions for the dative case. This map features a network of seven nodes which each represent a grammatical function. The Figure also illustrates that the English preposition *to* covers four of these functions (as was shown in example 8.20): purpose, direction, recipient and experiencer. It does not cover the functions beneficiary, predicative possessor (if prepositional verbs are not counted as in *the dog belongs to me*) and dative judicantis.

Semantic maps depend crucially on cross-linguistic research. For example, a node in the network is only added if at least one language is found which makes the distinction. Haspelmath gives the example of direction versus recipient (p. 217). Based on English and French, which use one preposition for both functions, this distinction could not be made. However, German uses *zu* or *nach* for direction, whereas it uses the dative case for recipient. A large sample set of languages is therefore needed to uncover all the uses of a gram.

Another important aspect of semantic maps is the connection between nodes in the network. The map must represent these nodes in a contiguous area on the map. Haspelmath writes that based on the English preposition *to*, for example, the following three orders could be possible for purpose, direction and recipient (p. 217, example 4):

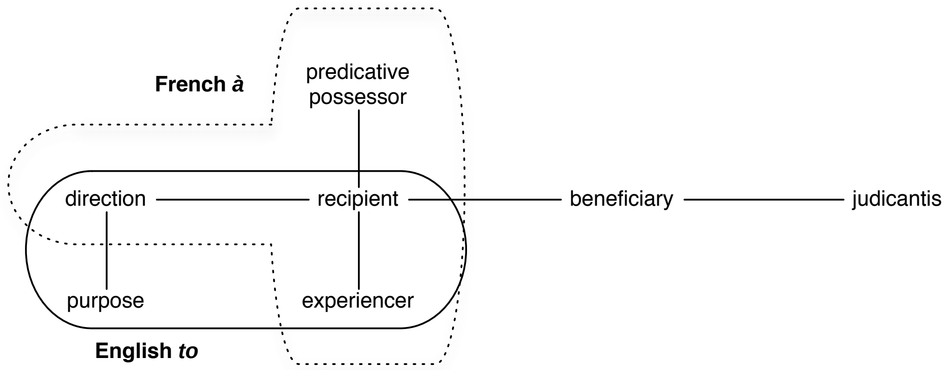


Figure 8.1: This partial semantic map compares the French preposition *à* to the English preposition *to* with respect to which typical dative functions they cover. Non-dative functions are ignored in this map (adapted from Haspelmath, 2003, figures 8.1 and 8.2, p. 213 and 215).

(8.23) a. purpose – direction – recipient
b. direction – purpose – recipient
c. direction – recipient – purpose

Again, data from other languages are taken into account for choosing which option can be eliminated. Since the French preposition *à* cannot be used for marking purpose, option (b) cannot represent a contiguous space in the network. The German preposition *zu* eliminates option (c) because it can express purpose and direction, but not recipient. The direct connections between functions on the semantic map are important because they are hypothesised to be universal:

> Semantic maps not only provide an easy way of formulating and visualizing differences and similarities between individual languages, but they can also be seen as a powerful tool for discovering universal semantic features that characterize the human language capacity. Once a semantic map has been tested on a sufficiently large number of languages [...] from different parts of the world, we can be reasonably confident that it will indeed turn out to be universal. (Haspelmath, 2003, p. 232)

This view is shared by many other linguists, among whom Bill Croft. Croft's *Semantic Map Connectivity Hypothesis* (Croft, 2001, p. 96) states that the functions of a particular construction will always cover functions that are connected regions in *conceptual space*. In other words, grammatical categories are language-particular, but they are based on a universal conceptual/semantic space.

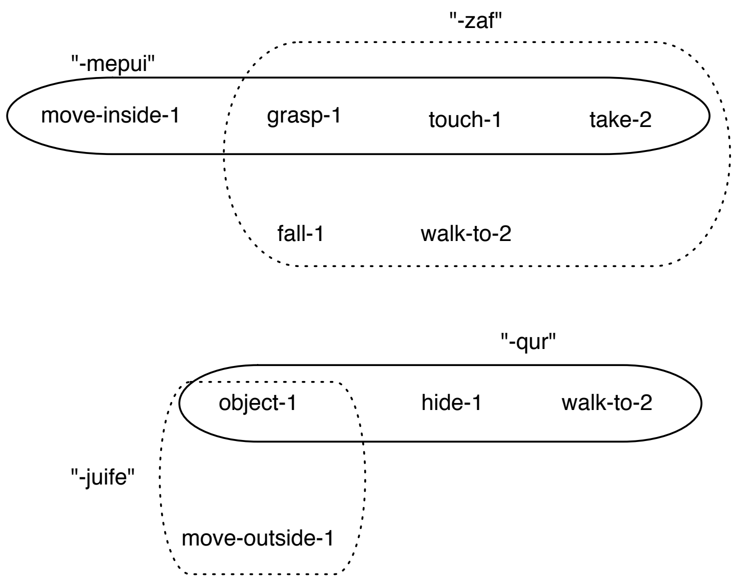


Figure 8.2: This diagram compares two different artificial grammars with respect to two categories in each of them. The languages were formed using the final set-up of experiment 3 (section 7.4). Even though the agents did not have a continuous conceptual space in advance, it is nevertheless possible to draw a primitive semantic map afterwards.

The universality of semantic maps is however an issue of debate. For example, Cysouw (in press) reports on his attempts at making a satisfying map for person marking. He concludes that there is no single 'universal' semantic map. Instead, different semantic maps are possible depending on the level and granularity of the analysis. Cysouw therefore calls for using semantic maps as a tool for modeling attested linguistic variety and as a way to predict *probable* languages rather than *possible* languages by weighting the function nodes in the network depending on the number of attested cases.

Cysouw thus points to a serious problem of the semantic map hypothesis: what grain-size is acceptable for making semantic maps? For example, Haspelmath (2003) uses functions such as 'recipient' and 'beneficiary' as primitive categories for his analysis. As I argued in Chapter 4, however, these functions are language-specific and no grammatical category has been demonstrated to cover all possible instantiations of such a function. Instead, languages tend to have many exceptions, irregularities or a redundant overlap in categories that mark that function. For example, 'recipient' and 'beneficiary' not only occur with the prepositions *to* and *for* respectively, they can also take the first object position in the English ditransitive.

The universality hypothesis therefore faces a problem of circularity. On the one hand, semantic maps are hypothesised to represent universal conceptual space; on the other hand, that conceptual space is based on an analysis which ignores language-internal differences and irregularities, and the languages that do not mark any differences are still assumed to have the same underlying functions.

A possible alternative. Artificial language evolution could demonstrate an alternative hypothesis to explain the universal tendencies in grammatical marking. In problem-solving models such as this thesis, grammatical evolution is a consequence of distributed processes whereby language users shape and reshape their language. The main challenge is therefore to find out what these processes are and under what circumstances they could create the kind of semantic maps that are observed for human languages. The hypothesis is that these processes suffice for the emergence of semantic maps and that conceptual space is dynamically configured in co-evolution with grammar. Semantic maps of different languages will naturally show similarities and differences depending on whether they followed the same evolutionary pathways or not.

Prior work on concept formation. As mentioned in section 3.3, prior work in the field has already demonstrated how a population of agents could self-organise a shared ontology through communicative interactions. Steels (1997a) reports the first experiments in which conceptualisation and lexicon formation are coupled to each other. In the experiment, a population of artificial agents take turns in playing ‘guessing games’: the speaker chooses one of the objects in the context to talk about and wants to draw the hearer’s attention to it by saying a word. The game is a success if the hearer points to the correct object. If the game fails, the speaker will point to the intended topic and the hearer tries to guess what the speaker might have meant with his word. The agents start without any language and even without an ontology. Instead, they are equipped with several sensory channels for perceiving the objects in their environment.

At the start of a game, two agents are randomly chosen from the population to act as a speaker and as a hearer. The speaker chooses an object from the context to talk about and needs to conceptualise a meaning which discriminates the topic from the other objects in the context. For example, if the topic is a green ball and there are also three red balls in the context, then the topic’s colour would be a good discriminating feature. At the beginning of the experiment, the agents have no concepts yet so the speaker has to create a new one. She will do so by taking the minimal set of features that can discriminate the topic from the other objects. The speaker will then invent a new word for this concept or meaning and transmit it to the hearer.

The hearer will in turn experience a communicative problem: she does not know the word that was used by the speaker. The game thus fails, but the speaker points to the intended object. The hearer then tries to retrieve the intended meaning through the same discrimination game. Often there are many different sets of discriminating features possible, but if the agents play a sufficient amount of language games with each other, they come to an agreement on what the form-meaning pairs are in their language and thus also reach a shared conceptual space. Similar experiments have been successfully performed in the domains of colour terms (Steels & Belpaeme, 2005) and spatial language (Steels & Loetzsch, 2008), and they have been scaled up to large meaning spaces (Wellens, 2008).

The contribution of this thesis. All of the above experiments confirm that communicative success can be a driving force for constructing an ontology of meaningful distinctions and that language can be used as a way to agree on a shared ontology among a population of autonomous embodied artificial agents. These experiments, however, have only focused on concept-and-lexicon formation so far, and the systematic relations between words have not been investigated yet. The experiments in this thesis, however, have polysemous semantic roles so they form an ideal starting point for testing the alternative hypothesis.

Figure 8.2 illustrates how analogical reasoning can be responsible for constructing coherent classes of semantic roles. The diagrams shows two semantic maps for two languages that were formed in the last set-up of experiment 3 as described in section 7.4 (multi-level selection with memory decay and pattern formation). Both diagrams show that it is possible to draw a primitive semantic map which compares the semantic roles of both languages. For example, in one language the marker *-mepui* can be used for covering four participant roles. Three of them (grasp-1, touch-1 and take-2) overlap with a semantic role of a different language. A similar observation counts for the two semantic roles in the second semantic map.

A comparison of the formed artificial languages suggests that grammaticalization processes can be visualised as a movement or change in connected regions of a continuous domain as a side-effect of analogical reasoning: extension of a category happens when new situations are encountered which are closely related to the existing categories. This shows that semantic maps could in principle be the result of dynamic processes involving analogy rather than starting from universal conceptual space.

The alternative proposed here needs further investigation and essentially requires a significant scale-up in terms of the meaning space and world environment as well as the conceptualisation capabilities of the agents. The present results are however encouraging and the proposed alternative has the advantage that it is more adaptive and open-ended to a changing environment: a universal conceptual space would still require some mechanism of mapping culture-specific developments (such as buying and selling, driving cars, and steering airplanes) onto a prewired structure. If the alternative hypothesis is followed, semantic maps would thus not point to a universal map of human cognition but rather to recurrent patterns in human experience and preferred developmental pathways followed by dynamic categorisation mechanisms.

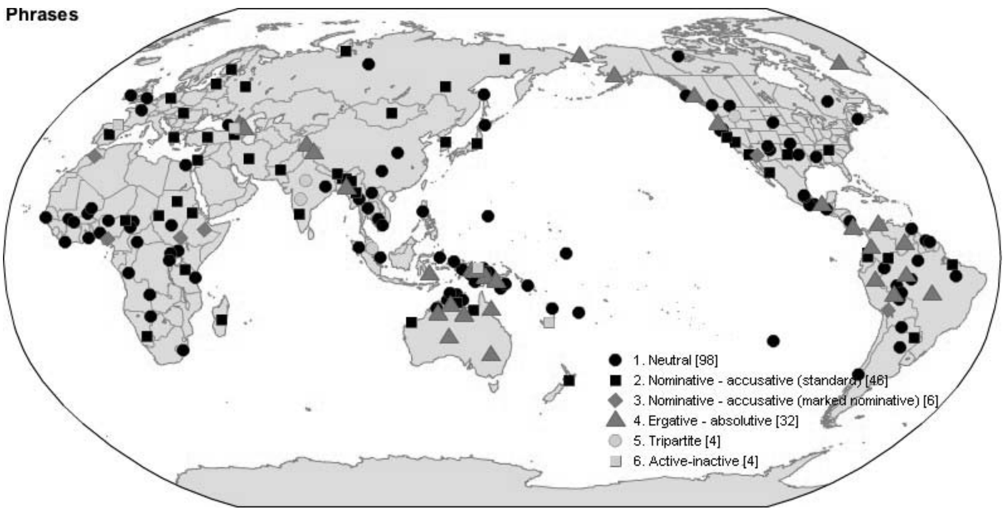


Figure 8.3: The alignment of case marking of full noun phrases (Comrie, 2005, 98).

8.4.2 Thematic hierarchies in case systems

In chapter 4 I discussed some linguistic evidence against the existence of universal thematic hierarchies both in the strong as in the implicational sense. Here, I will make a similar argument as I did for semantic maps and propose that preferences in argument linking can arise as a side-effect of analogical reasoning and multi-level selection.

The problem of thematic hierarchies and systematicity. One of the reasons for including a thematic hierarchy in linguistic theory was to account for alternating patterns of argument realisation such as *I opened the door with a key* versus *a key opened the door*. Another area in which thematic hiearchies thrived was linguistic typology as can be seen in the numerous proposals that have been made in the past (e.g. Dik, 1997; Fillmore, 1968; Givón, 2001; Jackendoff, 1990; Keenan & Comrie, 1977; Van Valin, 2004). Thematic hierarchies seemed very attractive to account for recurrent patterns of argument linking across languages and have therefore been hypothesised as a universal of human cognition. However, as the many proposals indicate and as I discussed in Chapter 4, thematic hierarchies can offer tendencies at best. Even for language-specific argument linking patterns, no satisfying hierarchy has been found yet.

The question of how language systematicity can ever arise becomes a big issue if no universal hierarchy can be found, especially if no Universal Grammar is assumed. The map in Figure 8.3, for example, shows the alignment of case marking of full noun phrases across 190 languages. It clearly demonstrates strong systematicity in the marking of 'core arguments' in these languages. Comrie (2005) distinguishes five different systems (I count the two variants of nominative-accusative systems as one):

- Neutral: the subject of intransitive clauses (S) is marked in the same way as both the subject (A) and object (P) of transitive clauses. Example: Mandarin.
- Nominative-accusative: A and S are marked in the same way (nominative marking). P is marked differently (accusative marking). Example: Latvian.
- Ergative-absolutive: S and P are marked in the same way (ergative marking), A is marked differently (absolutive marking). Example: Hunzib.
- Tripartite: S, A and P are all marked differently. Example: Hindi.
- Active-inactive: There is a different marker for an agentive S (aligning with A) and a patientive S (aligning with P). Example: Georgian.

The answer for most linguists is again sought in universals. For example, Croft (1998) assumes a universal conceptual space and universal linking rules for mapping arguments to core syntactic cases. The problem here is again that the proposals only work for analyses which do not go beyond the crude representation of case marking systems as presented in Figure 8.3. Closer studies show that the proposed systems are again only tendencies in each language and that there are lots of exceptions to the 'default' alignment of case marking. Also the typological variation across languages is greater than suggested by the traditional SAP-system of core arguments (Mithun, 2005).

Analogy, pattern formation and multi-level selection. In the case of thematic hierarchies, a similar alternative can be devised based on the distributed processes whereby language users shape and reshape their language for communication. As I argued in Chapter 6, generalisation of grammatical categories arises as a side-effect within inferential coding systems: language users want to increase their communicative success and when speakers have to solve a problem or innovate, they will try to do this in such a way that the intended communicative effect is still reached. By exploiting analogy, the speaker can hook the new situation up to previous conventions which are probably known by the hearer as well. The hearer can then retrieve the intended meaning through the same mechanisms of analogical reasoning.

As categories get reused more often, they increase their type frequency and hence their productivity. An additional factor that boosts the success of such a category is when it starts to form patterns or groups with other elements in the inventory. A multi-level selection alignment strategy then assures that certain categories can also survive and reoccur in multiple levels of the linguistic network which again increases their frequency and chances of survival. Multi-level selection could thus explain how different constructions align their categories with each other as demonstrated in the map in Figure 8.3.

Preferences in argument linking, as predicted by thematic hierarchies, could thus gradually emerge as a side-effect of these mechanisms: as certain categories become more and more dominant and productive, they can start to extend their use across patterns and eventually evolve into prototypical subject and object categories (as I also suggested in section 4.2.5). The many subregularities that are observed in languages are no problem in this model and are in fact predicted because everything has to emerge in a bottom-up fashion. Further experiments on the formation of syntactic cases could thus be the starting point for modeling this alternative to thematic hierarchies.

8.4.3 A redundant approach to grammaticalization

A third debate in which artificial language evolution can offer novel insights is grammaticalization theory. As I already mentioned in section 4.2.3, one of the problems of grammaticalization is that linguists can usually only detect language change once the processes of grammaticalization have already taken place. It is therefore difficult to hypothesise what mechanisms should be proposed to explain such changes especially since the consequences of communicative interactions in larger populations are often overlooked. Multi-agent simulations can thus demonstrate which mechanisms are better suited for dealing with innovations, variations, and propagations of linguistic conventions.

Reanalysis and actualization. Diachronic reanalysis has taken a foreground position in traditional grammaticalization theory. For example, Hopper & Traugott (1993) write: *"Unquestionably, reanalysis is the most important mechanism for grammaticalization"* (p. 32). Reanalysis is understood as a *"change in the structure of an expression or class of expressions that does not involve any immediate or intrinsic modification of its surface manifestation*" (Langacker, 1977, p. 59). In other words, reanalysis is not noticeable from the surface form but only has consequences for the grammar at a later stage. Many grammaticalization theories therefore posit another mechanism called 'actualization' which maps out the consequences of reanalysis (Timberlake, 1977).

Reanalysis is typically illustrated by the grammaticalization of *be going to* into *gonna* (Hopper & Traugott, 1993, p. 2–4). In an older use of *be going to*, *to* was part of a purposive directional complement as in *I am going to marry Bill* meaning 'I am going/travelling in order to marry Bill'. At a later stage, *to* is hypothesised to be reanalysed as belonging to *be going* instead of to the complement. In other words, rebracketing of the structure has taken place from [[I] [am going] [to marry Bill]] to [[I] [am going to] [marry Bill]].

The importance of reanalysis (and its combination with actualization) has recently been challenged. In section 2.3.1 I already argued that reanalysis essentially assigns language change to language acquisition, but that this model makes false predictions. Haspelmath (1998) adds that reanalysis does not entail a loss of autonomy which is typical for grammaticalization and that grammaticalization is (almost exclusively) unidirectional instead of bidirectional as predicted by reanalysis. Haspelmath also rejects

Grammaticalization	*Reanalysis*
loss of autonomy / substance	no loss of autonomy / substance
gradual	abrupt
unidirectional	bidirectional
no ambiguity	ambiguity in the input structure
due to language use	due to language acquisition

Table 8.3: This table shows the major differences between grammaticalization and reanalysis (Haspelmath, 1998, p. 327, Table 1).

the combination of reanalysis with actualization which is often used as a way to assign gradualness to reanalysis (p. 340–341). Actualization makes ‘reanalysis’ as a mechanism impossible to verify and it requires speakers to know at least two analyses of the same construction to account for both the old and the new behaviour. Actualization also does not explain how innovations might propagate. Haspelmath’s comparison of ‘grammaticalization’ and ‘reanalysis’ is summarised in Table 8.3.

An alternative for reanalysis. Despite the problems of reanalysis, it seems hard to conceive an alternative process that could explain certain changes. Haspelmath suggests that formal theories should implement the gradience of membership of word classes in some way such as *“$V_{1.0}$ for ordinary verbs, $V_{.7}$/$P_{.3}$ for preposition-like verbs (e.g.* considering*) and so on”* (p. 330). Even though gradience is indeed an important matter, such a proposal cannot capture the fact that the old use of a linguistic item and its new function can co-exist for hundreds of years in a language. The alternative that I would propose is redundancy and pattern formation along the lines of my example for French predicate negation in Chapter 7. Applied to the example of *gonna*, this alternative would simply state that the frequent co-occurrence of the words *be going to* led to the creation of a pattern for optimising linguistic processing. Once this pattern is created, it may start evolving on its own which allows it to gradually drift away from the original use of the words.

Example: the English verbal gerund. To illustrate this alternative approach, I will briefly take a look at the English verbal gerund which historically developed from a deverbal nominalisation and which later acquired more and more verbal properties. I will show examples of this development taken from Fanego (2004) and summarise how she describes this grammaticalization process in terms of ‘reanalysis’ and ‘actualization’. Next, I will argue for a simpler model based on redundancy and pattern formation.

The English gerund is a unique category in European languages in the sense that it is a third type of verbal complement besides to-infinitives (example 8.24) and finite clauses (8.25). The present-day English gerund has the following verbal properties: it can take a direct object (8.26), it can be modified by adverbs (8.27), it can mark tense, aspect and voice distinctions (8.28), it can be negated using the predicate negator *not* (8.29) and it can take a subject in a case other than the genitive (8.30).

(8.24) I just called *to say* 'I love you'.

(8.25) Just tell him *we're not interested anymore*.

(8.26) By writing *a book*, he managed to face all his inner demons.

(8.27) My *quietly* leaving before anyone noticed.

(8.28) The necessity of *being loved* is a driving force in our lives.

(8.29) My *not* leaving the room caused a stir.

(8.30) We should prevent *the treaty* taking effect.

Studies on the emergence and evolution of the gerund suggest that it developed from a deverbal nominalisation construction, similar to phrases such as *the writing of a book* (Tajima, 1985). This nominalisation lacked the aforementioned verbal properties, which can be illustrated with a similar nominalisation construction in Dutch: example 8.31 shows that the nominalized *bewerking* 'adaptation' cannot be complemented by a direct object (as is possible with the English gerund). Instead, it requires the genitival preposition *van* 'of' (a). Example (c) shows that speakers of Dutch need to combine a prepositional noun phrase with some kind of to-infinitive to express one of the functions carried by the English gerund.

(8.31) a. *de bewerking van het stuk*
the adaptation of the piece
'the adaptation of the play'

b. **door bewerking het stuk*
by adaptation the piece

c. *door het stuk te bewerken*
by the piece to adapt
'by adapting the play'

From this kind of deverbal nominalisation, the English gerund probably evolved according to the following steps (Tajima, 1985, summary and examples taken from Fanego, 2004):

1. Around 1200, the deverbal nominalisation *-ing* began taking adverbial modifiers of all kinds:

(8.32) *Of þi comyng at domesday*
'Of your coming *at doomsday*.'

2. The first examples with direct objects have been attested around 1300.

(8.33) *yn feblyng þe body with moche fastyng*
'in weakening the body by too much abstinence'

3. In the Early Modern English period, other verbal features are increasingly found, such as distinctions of voice and tense. From Late Modern English on, gerunds also start to take subjects:

(8.34) *he was war of hem comyng and of here malice*
'he was informed of them coming and of their wickedness'

Fanego (2004) argues that these changes are best understood as reanalysis of a nominal structure to a (more) verbal one (p. 26). This requires the speaker's ability to recognise multiple structural analyses since the 'old' and the 'new' use co-existed for a long time. The following examples show how the nominal analysis and the more verbal structure could be used together around 1300, whereas nowadays the nominal structure is unacceptable unless there is a determiner:

(8.35) *Sain Jon was ... bisi In ordaining of priestes, and clerkers, And in planning kirc werkes.*
'Saint John was ... busy ordaining priests and clerics, and in planning church works.'

(8.36) the ordaining of priests / the planning of works

(8.37) ordaining priests / planning works

(8.38) *ordaining of priests / *planning of works

In order to account for the gradualness of the change, Fanego suggests a reanalysis-plus-actualization model. She acknowledges Haspelmath (1998)'s criticism on this model that it is still not gradual enough and she proposes that the gerund should be regarded as a hybrid category which is partly noun and partly verb. To summarise, Fanego suggests that the development of the various uses of the English gerund involved (a) reanalysis and (b) actualization using Haspelmath's proposal for gradient categories.

Problems with Fanego's account. Fanego's analysis of the development of English requires complex cognitive operations from the part of the speaker that do not seem entirely justified. First of all, in order to reconcile reanalysis with the data, she needs to call on the process of actualization. However, as Haspelmath (1998) already noted, actualization *"waters down the notion of reanalysis, because it allows one to posit non-manifested reanalysis as one pleases"* (p. 341). It also seems contradictory to propose reanalysis, which is essentially an abrupt and discrete process, together with Haspelmath's gradient categories. Mechanisms such as semantic bleaching, analogy and extension could explain a gradual shift from a nominal category to a more verb-like category just as well without evoking reanalysis.

A second problem has to do with the idea of a gradient category, that is, analysing the gerund as some hybrid category which is let's say 20% nominal and 80% verbal. This kind of analysis treats the Gerund as a single category in the grammar whereas Fanego herself distinguishes at least three different types existing today, each with their own particular syntactic behaviours:

- Type 1: gerunds lacking determiners (e.g. *by writing it*)
- Type 2: gerunds taking determiners (e.g. *the writing of the letter*)
- Type 3: verbal gerund (e.g. *the people living in this town*)

A model based on redundancy. Reanalysis is a mechanism which is based on mismatches in learning. In the case of the English gerund, Fanego writes that the first gerunds to take verbal traits were the ones occurring in constructions without determiners (p. 19–20). However, the lack of determiners is in itself not necessarily a reason for reanalysing a grammatical structure, especially since determiners were not at all obligatory in many noun phrases in Old and Middle English (Traugott, 1992, p. 172–174). One can therefore reverse the question and ask why some uses of the gerund resisted the spread of determiners. In other words: is there a *functional* explanation for the development of the gerund?

A first step in the alternative hypothesis is to accept redundancy: language users store many instances in memory so rather than looking for a single category which leads to multiple structural analyses, speakers are assumed to store many instances in memory. Actual change in the system only takes place if one of these redundant instances gets extended. No layering or complex mechanisms for disambiguity are needed since there are still enough instances left that cover the older use of a particular form. Redundancy thus requires a far more simple cognitive model than the reanalysis-and-actualization approach and treats each use of the gerund as a construction in its own right.

Instead of reanalysis and mismatches in learning, the alternative hypothesis assumes that some patterns or instances extend their usage for a communicative reason. Fanego lists several possible sources (p. 11–17): first of all, the *-ing*-form of nominalisations was in competition with the Old English present participle *-ende* (which still exists in Dutch, for example). *-ing* became dominant by the fifteenth century and thus increased its frequency. Along with this competition, the productivity of *-ing* also increased from a limited number of verbs to an almost fully productive schema. A third possible source could be the fact that the English to-infinitive has resisted the combinations with other prepositions than *to* and *for to*. This created a gap in the usage of the infinitive which could be filled by the gerund (or conversely, the expansion of the gerund prevented the infinitive from filling this gap itself). Other sources are influences from French and the co-occurrence of the gerund with a genitive phrase.

The point here is not to find *the* source for the development of the English gerund but rather to illustrate that many possible sources can be identified and that they all probably played some role. It is therefore fruitful to see language as a selectionist system in which all linguistic items compete for a place in the inventory. Due to multi-level selection, categories can become more dominant across patterns which is what seemed to have happened with the gerund: it increased its productivity, won the competition against *-ende* for marking participles and hence became more frequent and successful.

Haspelmath (1998) also criticised reanalysis for failing to explain the strong unidirectional tendency of grammaticalization. In a system of multi-level selection, this could be explained due to the fact that once linguistic items become part of larger patterns or occur in multiple constructions, they are no longer fully independent of those constructions. The benefit of belonging to larger groups is that each item's individual survival chances increase, but the possible downside could be that the original use becomes structurally ambiguous or that it loses its distinctiveness. This would weaken its position and leaves the possibility for other items to conquer its space. In other words, there are always two factors influencing survival of a linguistic item: frequency and function.

Back to the computational model. The above analysis is only an illustration of how computational modeling could inspire linguists to come up with alternative hypotheses. The grammaticalization model of redundancy that I presented here mainly comes from the observation that variation in a population is an extremely challenging problem and that it is very difficult for a population to reach a shared and coherent language without losing generalisation accuracy. Moreover, the design stance can offer mechanisms and operationalisations that are simpler than the processes that are often proposed in verbal theories.

That said, the experiments presented in this thesis have not yet offered any proof that an analysis such as the one proposed here can actually work. However, they did show that a redundant and bottom-up approach *can* deal with high degrees of uncertainty in the development of a grammar whereas no such model exists (yet) for reanalysis. The comparison between the Iterated Learning Model (which essentially relies on reanalysis) and this thesis has shown that a usage-based approach performs significantly better than a reanalysis model. This does not mean that reanalysis does not exist or that it cannot be operationalised, but it poses some serious challenges to the effectiveness and explanatory power of the mechanism.

8.4.4 Summary: contributions of this thesis

In this section I advanced some novel hypotheses which contribute to the fields of linguistic typology and grammaticalization theory. Even though the experiments in this thesis are not conclusive enough to offer stronger support for these alternatives, the first results are nevertheless encouraging. I also suggested that artificial language evolution can demonstrate the consequences of the mechanisms that are proposed in linguistic debates and show which of these mechanisms are more efficient.

The first debate concerned the status of semantic maps as universals of human cognition. I pointed to literature in the field that gathers evidence against the universality of semantic maps and discussed some of the methodological and logical problems with semantic maps. No alternative hypothesis has really emerged, however, to explain the recurrent patterns of language development that are visualised by semantic maps. I suggested that semantic maps can emerge as a side-effect of analogy and multi-level selection as the result of distributed communicative interactions in populations of language users.

I then discussed a similar debate on the status of thematic hierarchies. The problems of thematic hierarchies are well-known in linguistic theory. Yet, they are still frequently used for describing recurrent argument realisation patterns across the world's languages which indicates that no viable alternative has propagated yet in the field. I argued that thematic hierarchies can again emerge as a side-effect of analogy and multi-level selection. Analogy can explain the increasing productivity of grammatical categories and multi-level selection can explain how systematicity rises across constructions.

Finally, I argued that artificial language evolution can provide highly relevant results for grammaticalization theory. Analysing the development of grammatical categories is notoriously difficult because there is a lack of first-hand empirical data that show innovations taking place. Linguists therefore have to fill in the blanks between old uses of a linguistic item and its new use. I suggested that artificial language evolution can demonstrate the effects of the mechanisms that are traditionally proposed to explain grammaticalization. I argued that the most widespread model of reanalysis-and-actualization requires a very complex cognitive apparatus in order to cope with variation in a population and proposed a simpler alternative based on redundancy, pattern formation and analogy.

8.5 Summary: impact and relevance

This chapter presented the possible impact of this thesis on artificial language evolution and discussed its contributions to linguistics. In the first section, I compared my experiments to a recent study on case marking in the Iterated Learning Model. The comparison showed that this thesis significantly advances the state-of-the-art in the field by presenting the first multi-agent simulations ever involving polysemous categories and by detecting and solving the problem of systematicity. I also argued that a cognitive-functional approach needs to be taken in order to move experiments on artificial language evolution towards greater complexity, expressiveness and realism.

The other three sections focused on the relevance of this thesis for linguistic theory. First of all, the representation of argument structure in this thesis offers the first computational formalism for argument realisation in a construction-based approach. I compared the Fluid Construction Grammar approach to an upcoming implementation in Sign-Based Construction Grammar, which is an extension of the more construction-based version of HPSG which has been developed during the past 10-15 years. One of the merits of my implementation as opposed to the SBCG model was that FCG does not require derivational constructions but that it allows more flexibility in argument realisation through its architecture of potential valents versus actual valency. I also argued against the notion of a 'minimal' lexical entry which goes against empirical data. Another salient feature of FCG is that it operationalises the more usage-based approach to language in a similar way as proposed in several current construction grammar theories such as Radical Construction Grammar, Cognitive Grammar and the Lakoff/Goldberg model.

Next, I discussed how construction grammar describes a speaker's knowledge of her language in terms of a structured network of linguistic items. I coupled the insights of the experiments to these proposals and argued that many proposed mechanisms such as analogy and type frequency can indeed play an essential part in the development of the structure of this inventory. I also argued that all current theories underestimate the problem of systematicity and that they need to incorporate multi-level selection in the form of systematicity links. I also suggested that further experiments are needed to see whether multi-level selection can be scaled up to more complex languages as a possible alternative to inheritance networks.

Finally, I discussed potential applications of artificial language evolution in the fields of language typology and grammaticalization. Even though the experiments in this thesis did not yield enough results yet for supporting these applications more strongly, they engendered novel ways of thinking about certain problems. In the case of linguistic typology, I provided alternative hypotheses for the emergence of semantic maps and for tendencies in argument linking patterns. I argued that a combination of analogical reasoning and multi-level selection could yield the kind of universal preferences that theories of 'conceptual space' and thematic hierarchies try to capture. In the debate on grammaticalization and reanalysis, I took the position that reanalysis requires a very complex cognitive model and that a redundancy-based approach involving analogy and multi-level selection could offer an alternative hypothesis.

The results obtained in this thesis can obviously never offer any conclusive evidence for all of the aforementioned issues. There are many questions that the methodology cannot address and other disciplines should all provide different pieces of the puzzle to reach a coherent theory about the language faculty. In order to make further progress, a significant scale-up also needs to be made in terms of population size, meaning space and world environment. The present experimental results are however encouraging and a comparison to other approaches in the field of artificial language evolution strongly suggest that the right path is being explored. This means that in the future artificial language evolution can grow further and become increasingly important as a way to model and test verbal theories of language.

Chapter 9

Conclusions and Further Work

This thesis investigated how a case marking system can be developed as the consequence of locally situated interactions in a population of autonomous artificial agents that shape and reshape their language in order to optimise their communicative success. Its main objective was to move prior work on concept and lexicon formation into the domain of grammar and to push the state-of-the-art in research on artificial language evolution. Throughout this thesis, special attention was also given to its second objective: to engage in a real two-way dialogue with linguistics and to help creating the conceptual foundations that are necessary for communicating the results of one field to the other.

9.1 Summary

Communication, communication, communication. This mantra seems to make up the fabric of this thesis. It started in Chapter 2 in which I defended a cognitive-functional approach against the dominant but crumbling tradition of generative linguistics. In that chapter, I argued that language and grammar primarily serve communicative purposes in social interactions. I illustrated the general approach to grammar that is underlying this thesis and introduced various threads which would reappear throughout this work and make connections with other threads.

The first connections between artificial language evolution and linguistics were made in Chapter 3. Here I gave an overview of the three main approaches in the field of the origins and evolution of language and explained how they relate to the linguistic debates discussed in Chapter 2. I also discussed the methodology that I implemented in this thesis and gave a historical background of the prior work on top of which my research is built. This thesis took a design stance towards the formation of grammar but committed itself to aim for relevance in cognitive science as well.

Chapter 4 completed Part I and therefore the theoretical foundations of this dissertation. This chapter offered a more domain-specific overview of the problem of argument realisation in linguistics and some empirical data concerning the development of case markers in natural languages. As I argued later in this thesis, it is important to have a good appreciation of the complexity that is involved in grammatical phenomena. A clear understanding of what is known about case markers has helped to build a roadmap for studying the formation of case grammars and has prevented hasty conclusions of experimental results.

The second part started immediately with a technical chapter on argument realisation in Fluid Construction Grammar. This chapter provided an introductory sketch of FCG and exemplified the representation of argument structure constructions that was used in the experiments. This representation is the first construction-based approach of argument structure constructions to be implemented in a computational formalism.

Chapter 6 introduced the first series of experiments. After a successful replication of two-agent simulations performed by Luc Steels, this chapter reported a scale-up of the experiments towards a multi-agent population. Since the phrase *two's company, three's a crowd* works particularly well for multi-agent simulations, solutions had to be found for exploiting the full generalisation power of analogy and for allowing the agents to align polysemous semantic roles with each other. This chapter featured a redundant, bottom-up approach to the formation of grammar in which extension by analogy was accompanied by careful abstraction.

In the second series of experiments, agents were given the capacity to combine existing case markers into larger argument structure constructions through pattern formation. The experimental results revealed that languages become unsystematic once the linguistic inventory contains multiple hierarchical levels and if the inventory is unstructured. Further experiments demonstrated that an alignment strategy based on multi-level selection can solve this problem. Chapter 7 finished Part II with some first steps towards the formation of syntactic cases and some ideas for future research.

Finally, all the threads of this thesis came together again in Chapter 8 in which I discussed the impact of this thesis on the field of artificial language evolution and its contributions to linguistics. A comparison between my work and another study on the formation of case marking showed that this thesis makes significant progress in the domain of artificial language evolution. The chapter also illustrated how the experiments make highly relevant contributions to some important issues in linguistics such as the formalisation of argument structure in construction grammar, the structure of the linguistic inventory, the status of semantic maps and thematic hierarchies, and plausible mechanisms for explaining grammaticalization.

9.2 Advances of the research

Advances in artificial language evolution

This thesis reported two innovating experiments in the field of artificial language evolution. The first series of experiments offered the first multi-agent simulations ever which involved polysemous categories in the form of semantic roles. The second series of experiments identified and solved the problem of systematicity. I also showed that this problem has occurred in other experiments as well but that it was never noticed before. This led to misinterpretation of experimental results in the past.

The experiments demonstrated an operational set-up in which autonomous artificial agents were able to self-organise a shared grammatical system containing semantic roles. This grammatical system also included the combination of markers into larger argument structure constructions and the possibility of using words for events in multiple argument realisation patterns. The experiments succeeded in identifying and demonstrating the cognitive mechanisms and external pressures that were minimally required to make the transition of lexical languages to the grammatical systems developed in the simulations.

With respect to the external and communicative pressures, the agents had to talk about dynamic events to each other. This created a cognitive load for the hearer if the agents only had a lexical language. The agents were therefore endowed with an architecture of diagnostics for detecting communicative problems and opportunities for optimisation, repair strategies for solving these problems, and alignment strategies for coordinating their linguistic inventories.

This thesis mainly focused on the combination of analogy and multi-level selection. Analogy was demonstrated to be a powerful mechanism which can explain an increased type frequency and productivity of grammatical categories. The generalisation rate of analogy can only be maintained, however, if the agents not only innovate in a systematic way but also display systematic behaviour in consolidation. This thesis therefore proposed multi-level selection, an alignment strategy which favours the selection of groups of systematically related constructions.

In sum, the work in this thesis provides a significant step forward in the field of artificial language evolution. A comparison with a similar study in the Iterated Learning Model also showed that its methodology and approach is a more fruitful way of investigating the formation of grammar.

Relevance for linguistics

The experiments in this thesis required some technical and conceptual innovations which are highly relevant for linguistics as well. First of all, an efficient representation had to be found which could handle the variety and uncertainty of argument realisation patterns in the formation of grammar. The solution presented in this thesis implements the ‘fusion’ of verb-specific participant roles with semantic roles of argument structure constructions. This fusion process has so far only been described in terms of vague and contradictory principles and some recent developments in construction grammar such as SBCG even take a step back towards more traditional lexicalist accounts.

The representation introduces some novel ideas into linguistics. It rejects the notion of a minimal lexical entry but allows lexical items to list their potential valents. The semantics of the verb are therefore still crucial in determining its morpho-syntactic behaviour. However, its actual valency is also dependent on the constructions that combine with the verb. The architecture of potential valents versus actual valency could be expanded to the valency of other linguistic categories as well such as the various kinds of noun phrases in which a noun can occur. In this way, syntactic constraints do not have to be posited in one fixed place in the grammar but can be distributed and be dependent on the combinations in the utterance. This also gives the language user more freedom in shaping her language as claimed by usage-based models.

The combination of analogy for innovation and multi-level selection for consolidation that was applied in the experiments makes a significant contribution to many important issues in linguistics. I argued that linguistic theories have overlooked the importance of multi-level selection and that they focus too strongly on top-down instance links. I also discussed problems with theories on universal conceptual space, thematic hierarchies and reanalysis for grammaticalization and argued that no viable alternatives have propagated yet in the community. In each debate, I suggested that analogy and multi-level selection could solve many of the issues, but that additional experiments are needed in order to demonstrate the true power of these mechanisms.

This thesis thus succeeds in generating interesting insights for linguistics. Whether its second objective – establishing the first conceptual foundations for communicating results from one field to another – is reached, however, cannot be assessed here. This will depend on how this thesis is received in both communities. At least it is one of the first studies to offer a more profound coupling of the fields, which may inspire other researchers to do the same in future work.

9.3 Future work

In Part II of this thesis I already suggested which directions need to be followed for future work. I will summarise the most important issues here.

First of all, the experiments in this thesis scaffold the recruitment of a lexical item in Stage 2 of the development of case markers. As I argued in Chapter 6, using analogy all the way would optimise communication even further because it provides the hearer with a stronger grounding point for retrieving the meaning or function of a grammatical form. Implementing this stage is however a very complex matter which probably requires research on the coordination of syntactic structures and ellipsis of shared arguments. I therefore suggested that more research is needed to understand both how this process happens in natural languages and how syntactic categories can be formed which can be coordinated.

In Chapter 7 I provided the first steps towards such syntactic structures but I demonstrated that the agents currently do not get enough feedback in order to align an abstract and intermediate layer of semantic and syntactic categories. On a conceptual level, we therefore need to investigate how causal, aspectual and temporal distinctions can become relevant for communication and how this may lead to alternating argument structures. On the grammatical level, we need to have a more flexible analogy algorithm for allowing semantic roles to expand their use even further and to start taking over the functions of other roles. The analogy algorithm also needs to be adapted so that it can include patterns in its search.

Another major point of investigation concerns the repair strategies implemented in this thesis. I explicitly used the word ‘formation’ instead of ‘emergence’ of case grammars because the agents have no other choice but to form a case-like grammar. In other words, the repair strategies that I implemented were already domain-specific. One of the most interesting challenges of this line of work, I believe, is therefore to find out how more general-purpose cognitive mechanisms can be recruited and specialised by the agents themselves in order to create some kind of meta-rule or template for innovation. I strongly believe that a combination of analogy and pattern formation could play an important role in solving this issue.

Finally, as for all simulations in the field, future research has to scale up the experiment on all levels: a larger world and meaning space, larger populations, an integrated experiment involving also the formation of a lexicon, and all the other dimensions that need to be taken into account for testing the validity and robustness of the model.

9.4 And finally...

And finally I arrive at the last sentences of this dissertation. While writing these words, I am experiencing several contradictory feelings that are undoubtedly shared by all researchers over the world. I feel a sense of satisfaction because I discovered the beauty of case marking systems and a sense of hope that this thesis has helped to reach a deeper understanding of them. But at the same time I am thirsty for more and I feel the frustration of not knowing how the story ends. I feel as if I have just finished part one of an intriguing saga and that grammar has much more than just a few more tricks up its sleeve to surprise me...

Be as it may, the time has come to write the concluding sentence. Having written most parts of this thesis in Paris, it is only normal that I end this work by quoting the famous French linguist Antoine Meillet. He probably had a more structuralist interpretation in mind when he wrote down these words, but I recently rediscovered them in the light of my own work on analogy and multi-level selection: *"chaque langue forme un système où tout se tient."*

Appendices

Appendix A

Measures

The simulations reported in this thesis make use of a number of measures for assessing the progress made during the experiments. This appendix collects and explains them all both in order to provide the reader with a clear understanding of what is being measured and in order to provide the research community with clear definitions of measures for future experiments.

A.1 Communicative success

Communicative success as a local measure. Communicative success can be measured by the agents themselves and it can influence their linguistic behaviour. In the description games played in the experiments in this thesis, a game is a success if the hearer signals agreement with the speaker's description and a failure if the hearer signals disagreement. The hearer will agree if interpretation yields a single set of bindings between the parsed meaning and the facts in the memory. The hearer will disagree if interpretation is ambiguous (i.e. more than one hypothesis was returned) or if interpretation failed.

Plotting communicative success. Communicative success can also be plotted for a series of interactions by recording the success or failure of every language game. This is a global measure which is not observable by the agents themselves and thus has no influence on their linguistic behaviour. Each successful game is counted as 1 and each failed game is counted as 0. The sum of these results is then divided by the size of a certain interval into a single number between 0 and 1. The interval in all the reported simulations is set to 10.

$$\text{Result of game}_i = \begin{cases} 1 & \text{if game}_i \text{is successful} \\ 0 & \text{if game}_i \text{is successful} \end{cases}$$

$$\text{Communicative success } \begin{matrix} n \\ m \end{matrix} = \frac{1}{(n-m)} \sum_{i=m}^{n} \text{Result of game}_i$$

A.2 Cognitive effort

Cognitive effort as a local measure. Local cognitive effort is defined in this thesis as the number of inferences the hearer has to make during interpretation (i.e. the number of variables that need to be made equal). Since the event types in the simulations take a maximum of three participant roles, this measure ranges from 0 to 3. This number is recalculated onto a scale between 0 and 1 by taking the effort and dividing it by the maximum number of inferences (which is 3). One inference thus returns 0.33, two inferences 0.66 and three inferences 1. Failed language games count as 1, which is the maximum effort score. The agents use cognitive effort as one of the triggers for expanding their language.

Plotting cognitive effort. Cognitive effort can be plotted for a series of interactions by recording the hearer's effort during each interaction. Again, this is a global measure which is only accessible for the experimenter but not for the agents themselves. As with communicative success, cognitive effort in each game returns a value between 0 and 1. Global cognitive effort is measured by dividing the sum of the results by the size of a certain interval (which here is 10). This returns a measure between 0 and 1. In the following formulae CG_i stands for 'the hearer's cognitive effort during game$_i$'.

$$\text{CG}_i = \begin{cases} \text{Success}_i & = \dfrac{\text{number of inferences}}{\text{maximum number of inferences}} \\ \text{Failure}_i & = 1 \end{cases}$$

$$\text{Cognitive effort}\ {}^{n}_{m} = \frac{1}{(n-m)} \sum_{i=m}^{n} \text{CG}_i$$

A.3 Average preferred lexicon

The average preferred lexicon is used by various measures in this thesis. This lexicon is derived by taking the most frequent form for every possible meaning in the population. For example, if six agents in a population of ten prefer the marker *-bo* for the participant role 'move-1' as opposed to four agents that prefer the marker *-ka*, then *-bo* is listed in the average preferred lexicon with a frequency of 0.6. This lexicon is calculated for each individual participant role and for each possible combination of participant roles. For the experiments in this thesis, the complete meaning space of participant roles consists of the following meanings (of which the numbers correspond to the numbers in Figures 7.7, 7.8, 7.14 and 7.15):

1. object-1
2. move-1
3. visible-1
4. approach-1
5. approach-2
6. distance-decreasing-1
7. distance-decreasing-2
8. fall-1
9. fall-2
10. grasp-1
11. grasp-2
12. hide-1
13. hide-2
14. move-inside-1
15. move-inside-2
16. move-outside-1
17. move-outside-1
18. touch-1
19. touch-2
20. walk-to-1
21. walk-to-2
22. cause-move-on-1
23. cause-move-on-2
24. cause-move-on-3
25. give-1
26. give-2
27. give-3
28. take-1
29. take-2
30. take-3
31. approach-1 approach-2
32. distance-decreasing-1 distance-decreasing-2
33. fall-1 fall-2
34. grasp-1 grasp-2
35. hide-1 hide-2
36. move-inside-1 move-inside-2
37. move-outside-1 move-outside-2
38. touch-1 touch-2
39. walk-to-1 walk-to-2
40. cause-move-on-1 cause-move-on-2
41. cause-move-on-1 cause-move-on-3
42. cause-move-on-2 cause-move-on-3
43. give-1 give-2
44. give-1 give-3
45. give-2 give-3
46. take-1 take-2
47. take-1 take-3
48. take-2 take-3
49. cause-move-on-1 cause-move-on-2 cause-move-on-3
50. give-1 give-2 give-3
51. take-1 take-2 take-3

A.4 Meaning-form coherence

Meaning-form coherence is a global measure which is not accessible to the agents. It takes the most frequent form for a particular meaning (i.e. the form which is preferred by most agents in the population) from the average preferred lexicon. For example, if the marker *-bo* is preferred by six agents in a population of ten agents, it is listed in the preferred average lexicon with a frequency score of 0.6. Meaning-form coherence calculates the average of all these individual frequency scores:

$$\text{MF coherence} = \frac{\text{sum of all frequency scores in preferred average lexicon}}{\text{number of entries in preferred average lexicon}}$$

A.5 Systematicity

Systematicity is again a global measure which is not accessible to the agents. It is calculated by taking each meaning in the average preferred lexicon and comparing it to the combinations of meanings in which it occurs. If the combination uses the same marker as the relevant meaning, then a score of 1 is counted. If it is not, a score of 0 is counted. The sum of all these scores is divided by the number of meanings that had to be checked in the average lexicon, which yields a score between 0 (no systematicity) and 1 (maximum systematicity).

For example, suppose that the meaning 'appear-1' is most frequently marked by *-bo*, 'appear-2' by *-ka* and the combination of the two as *-bo -si*. First we take 'appear-1' and check whether its marker also occurs in the combination with appear-2: this is indeed the case so the form-meaning mapping is systematic in both constructions, which is counted as '1'. For appear-2, however, the pattern uses a different marker *-si* so no systematic relation exists across patterns, which is counted as 0. The combination itself does not occur in a larger pattern so it is not considered by the systematicity measure.

Appendix B

Abbreviations

1	first person
2	second person
3	third person
A	subject of transitive clauses
ACC	accusative
AM	Analogical Modeling
BCG	Berkeley Construction Grammar
BEN	benefactive
CAS	complex adaptive system
CG	Cognitive Grammar
DAT	dative
DECL	declarative
DET	determiner
ECG	Embodied Construction Grammar
F	feminine
FCG	Fluid Construction Grammar
INGR	ingressive
ILM	Iterated Learning Model
INFL	inflection / inflected form
LS	logical structure (in RRG)
M	masculine
MBLP	Memory-Based Language Processing
NEG	negation
NEW TOP	new topic
NLP	Natural Language Processing
NOM	nominative
OED	Old English Dictionary
Obj	object

P	object of transitive clauses
PAST	simple past tense
PL	plural
P&P	Principles & Parameters
pos	part of speech
PRES	simple present tense
RCG	Radical Construction Grammar
RRG	Role & Reference Grammar
S	subject of intransitive clauses
SBCG	Sign-Based Construction Grammar
Sbj	subject
SG	singular
SVO	subject-verb-object
TOP	topic
UG	Universal Grammar
V	verb

Index

References

ABRAMS, M. & GREENBLATT, S., eds. (2000). *The Norton Anthology of English Literature*, vol. 1. Norton, London, 7th edn. 36

ANDERSEN, H. (1973). Abductive and deductive change. *Language*, **49**, 765–793. 21

BAILLIE, J.C. (2002). *Apprentissage et Reconnaissance d'Actions dans des Sequences Video*. Ph.D. thesis, Universite de Paris. 131

BAILLIE, J.C. & GANASCIA, J.G. (2000). Action categorization from video sequences. In W. Horn, ed., *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI)*, 643–647, IOS Press, Berlin, Germany. 131

BALDWIN, J.M. (1896). A new factor in evolution. *American Naturalist*, **30**, 441–451. 30

BARLOW, M. & KEMMER, S., eds. (2000). *Usage-Based Models of Language*. Chicago University Press, Chicago. 28

BARONCHELLI, A., FELICI, M., LORETO, V., CAGLIOTI, E. & STEELS, L. (2006). Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics*, **P06014**. 52, 241

BATALI, J. (1998). Computational simulations of the emergence of grammar. In J.R. Hurford, M. Studdert-Kennedy & C. Knight, eds., *Approaches to the Evolution of Language: Social and Cognitive Bases*, 405–426, Edinburgh University Press, Edinburgh. 51

BATALI, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In T. Briscoe, ed., *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, Cambridge University Press, Cambridge. xviii, 197, 199

BELPAEME, T. (2002). *Factors Influencing the Origins of Colour Categories*. Ph.D. thesis, Vrije Universiteit Brussel. 52

BELPAEME, T. & BLEYS, J. (2005). Colourful language and colour categories. In C. Lyon, A. Cangelosi & C.L. Nehaniv, eds., *AISB'05: Proceedings of Second International Symposium on the Emergence and Evolution of Linguistic Communication (EELC'05)*, Hatfield, UK. 52

BERGEN, B.K. & CHANG, N. (2005). Embodied Construction Grammar. In J.O. Östman & M. Fried, eds., *Construction Grammars: Cognitive Grounding and Theoretical Extensions*, John Benjamins, Amsterdam. 95

BLAKE, B.J. (1994). *Case*. Cambridge Textbook in Linguistics, Cambridge University Press, Cambridge. 65, 77, 80, 81, 88, 91, 179, 229

BLEYS, J. (2008). The origins of recursive rules: A usage-based account. In A.D. Smith, K. Smith & R. Ferrer i Cancho, eds., *The Evolution of Language. Proceedings of the 7th International Conference (EVOLANG 7)*, 34–41, World Scientific Press, Singapore. 54, 55

BLOOMFIELD, L. (1970). *Language*. Allen and Unwin, London, originally published in 1933. 12

BRESNAN, J. (1982). *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge, MA. 69, 95

BRIGHTON, H., KIRBY, S. & SMITH, K. (2005a). Cultural selection for learnability: Three principles underlying the view that language adapts to be learnable. In M. Tallerman, ed., *Language Origins: Perspectives on Evolution*, Chapter 13, Oxford University Press, Oxford. 39, 40, 236

BRIGHTON, H., KIRBY, S. & SMITH, K. (2005b). Language as an evolutionary system. *Physics of Life Reviews*, **2**, 177–226. 39, 41

BRISCOE, E. (2000). Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language*, **76**, 245–296. 36

BYBEE, J. (1998). A functionalist approach to grammar. *Evolution of Communication*, **2**, 249–278. 20

BYBEE, J. & THOMPSON, S. (2000). Three frequency effects in syntax. In *Berkeley Linguistics Society*, vol. 23, 65–85, University of California, Berkeley, Dwinelle Hall, Berkeley CA. 161

CHAFE, W., ed. (1980). *The Pear Stories*. Ablex, New York. 82

CHOMSKY, N. (1957). *Syntactic Structures*. Mouton, The Hague. 25, 26

CHOMSKY, N. (1959). A review of B.F. Skinner's Verbal Behavior. *Language*, **35**, 26–58. 16

CHOMSKY, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA. 12, 16, 26

CHOMSKY, N. (1980). *Rules and Representations*. Basil Blackwell, Oxford. 13

CHOMSKY, N. (1986). *Knowledge of Language: Its Nature, Origins, and Use*. Praeger, New York. 16, 17, 21

CHOMSKY, N. (1995). *The Minimalist Program*. MIT Press, Cambridge, MA. 13, 31

CHOMSKY, N. (2005). Three factors in language design. *Linguistic Inquiry*, **36**, 1–22. 13

CHOMSKY, N. & LASNIK, H. (1993). Principles and parameters theory. In J. Jacobs, A. von Stechow, W. Sternefelt & T. Vennemann, eds., *Syntax: An International Handbook of Contemporary Research*, vol. 1, 506–569, Mouton de Gruyter, Berlin. 18

CHRISTIANSEN, M. & KIRBY, S. (2003). Language evolution: The hardest problem in science? In M. Christiansen & S. Kirby, eds., *Language Evolution*, 1–15, Oxford University Press, Oxford. 45

COMRIE, B. (2005). Alignment of case marking. In M. Haspelmath, M.S. Dryer, D. Gil & B. Comrie, eds., *The World Atlas of Language Structures*, chap. 98–99, Oxford University Press, Oxford. xviii, 268

CROFT, W. (1991). *Syntactic Categories and Grammatical Relations. The Cognitive Organization of Information*. Chicago UP, Chicago. 19, 58, 59, 60, 65, 67, 68, 82, 87

CROFT, W. (1998). Event structure in argument linking. In M. Butt & W. Geuder, eds., *The Projection of Arguments: Lexical and Compositional Factors*, 21–63, CSLI Publications, Stanford. 63, 64, 65, 66, 67, 68, 133, 269

CROFT, W. (2000). *Explaining Language Change: An Evolutionary Approach*. Longman, Harlow Essex. 22, 23, 24, 42

CROFT, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford UP, Oxford. 19, 28, 60, 65, 69, 76, 256, 264

CROFT, W. (2003). Lexical rules vs. constructions: A false dichotomy. In H. Cuyckens, T. Berg, R. Dirven & K.U. Panther, eds., *Motivation in Language Studies: Studies in Honour of Günter Radden*, 49–68, John Benjamins, Amsterdam. 70, 72, 74, 75, 97

CROFT, W. (2004). The relevance of an evolutionary model to historical linguistics. In O. Nedergaard Thomsen, ed., *Different Models of Linguistic Change*, John Benjamins, Amsterdam. 42

CROFT, W. (2005). Logical and typological arguments for radical construction grammar. In J.O. Östman & M. Fried, eds., *Construction Grammars: Cognitive Grounding and Theoretical Extensions*, 273–314, John Benjamins, Amsterdam. 54

CROFT, W. (2008). The origins of grammaticalization in the verbalization of experience. *Linguistics*, to appear. 80, 82

CROFT, W. (in prep.). *Verbs: Aspect and Argument Structure*. Oxford University Press, Oxford, manuscript consulted on 16 October 2006. 67, 68

CROFT, W. & CRUSE, D.A. (2004). *Cognitive Linguistics*. Cambridge Textbooks in Linguistics, Cambridge University Press, Cambridge. 43, 184, 255, 256, 258, 259

CYSOUW, M. (in press). Building semantic maps: The case of person marking. In M. Miestamo & B. Wälchli, eds., *New Challenges in Typology*, Mouton De Gruyter, Berlin. 265

DAELEMANS, W. (2002). A comparison of Analogical Modeling of Language to Memory-Based Language Processing. In R. Skousen, D. Lonsdale & D. Parkinson, eds., *Analogical Modeling*, 157–179, John Benjamins, Amsterdam. 220

DAELEMANS, W. & VAN DEN BOSCH, A. (2005). *Memory-Based Language Processing*. Studies in Natural Language Processing, Cambridge University Press, Cambridge. 7, 27, 28, 29, 164, 219, 220, 260

DAELEMANS, W., VAN DEN BOSCH, A. & ZAVREL, J. (1999). Forgetting exceptions is harmful in language learning. *Machine Learning*, **34**, 11–43. 29

DAVIDSE, K. (1996). Functional dimensions of the dative in English. In W. Van Belle & W. Van Langendonck, eds., *The Dative. Volume 1: Descriptive Studies*, Case and Grammatical Relations Across Languages, 289 – 338, John Benjamins, Amsterdam. 60, 72

DAVIES, W.D. (1986). *Choctaw Verb Agreement and Universal Grammar*. Reidel Publishing Company, Dordrecht. 90

DAVIS, A. & KOENIG, J.P. (2000). Linking as constraints on word classes in a hierarchical lexicon. *Language*, **76**, 59–91. 253

DE BEULE, J. (2007). *Compositionality, Hierarchy and Recursion in Language. A Case Study in Fluid Construction Grammar*. Ph.D. thesis, Vrije Universiteit Brussel. 43, 54, 99, 105, 200

DE BEULE, J. (2008). The emergence of compositionality, hierarchy and recursion in peer-to-peer interactions. In A.D. Smith, K. Smith & R. Ferrer i Cancho, eds., *The Evolution of Language. Proceedings of the 7th International Conference (EVOLANG 7)*, 75–82, World Scientific Press, Singapore. 43, 54, 246

DE BEULE, J. & BERGEN, B.K. (2006). On the emergence of compositionality. In A. Cangelosi, A. Smith & K. Smith, eds., *The Evolution of Language*, World Scientific, Singapore. 42, 43, 44, 54, 200

DE BEULE, J. & STEELS, L. (2005). Hierarchy in Fluid Construction Grammar. In U. Furbach, ed., *KI 2005: Advances In Artificial Intelligence. Proceedings of the 28th German Conference on AI*, vol. 3698 of *Lecture Notes in Artificial Intelligence*, Springer, Koblenz. 54, 105

DE BOER, B. (1999). *Self-organisation in Vowel Systems*. Ph.D. thesis, Vrije Universiteit Brussel, Brussels. 238

DE BOER, B. (2000). Self organization in vowel systems. *Journal of Phonetics*, **28**, 441–465. 47, 55

DE PAUW, G. (2002). *An Agent-Based Evolutionary Computing Approach to Memory-Based Syntactic Parsing of Natural Language*. Ph.D. thesis, Universiteit van Antwerpen, Antwerp. xviii, 135, 198, 199

DE SAUSSURE, F. (1916). *Cours de Linguistique Générale*. Payot, Paris, sixth, revised, 2005 edn., Édition Critique Préparée par Tulio de Mauro, first published in 1967. 24, 25

DE VYLDER, B. (2007). *The Evolution of Conventions in Multi-Agent Systems*. Ph.D. thesis, Vrije Universiteit Brussel, Brussels. 52, 152, 241

DIESSEL, H. (2005). Distance contrasts in demonstratives. In M. Haspelmath, M.S. Dryer, D. Gil & B. Comrie, eds., *The World Atlas of Language Structures*, 170–173, Oxford University Press, Oxford. 32

DIESSEL, H. & TOMASELLO, M. (2000). The development of relative constructions in early child speech. *Cognitive Linguistics*, **11**, 131–152. 18

DIK, S. (1997). *The Theory of Functional Grammar*. Mouton de Gruyter, Berlin. 13, 14, 15, 16, 20, 23, 64, 69, 268

DOWTY, D. (1979). *Word Meaning and Montague Grammar*. Reidel, Dordrecht. 61

DOWTY, D. (1991). Thematic proto-roles and argument selection. *Language*, **67**, 547–619. 63, 64

DRYER, M.S. (1997). Are grammatical relations universal? In J. Bybee, J. Haiman & S. Thompson, eds., *Essays on Language Function and Language Type: Dedicated to T. Givon*, 115–143, John Benjamins, Amsterdam. 19, 20, 60

FANEGO, T. (2004). On reanalysis and actualization in syntactic change: The rise and development of English verbal gerunds. *Diachronica*, **21**, 5–55. 271, 272, 273

FILLMORE, C. (1968). The case for case. In E. Bach & R. Harms, eds., *Universals in Linguistic Theory*, 1–88, Holt, Rhinehart and Winston, New York. 58, 62, 63, 64, 69, 268

FILLMORE, C. (1970). The grammar of hitting and breaking. In R. Jacobs & P. Rosenbaum, eds., *Readings in English Transformational Grammar*, 120–133, Ginn, Waltham, MA. 26, 58

FILLMORE, C.J., KAY, P. & O'CONNOR, M.K. (1988). Regularity and idiomaticity in grammmatical constructions: the case of *let alone*. *Language*, **64**, 501–538. 27

FILLMORE, C.J., KAY, P., MICHAELIS, L. & SAG, I. (in preparation). *Construction Grammar*. Chicago University Press, Chicago, unpublished manuscript. Chapter 7 available online at http://lingo.stanford.edu/sag/SBCG/7.pdf. Last accessed on 1 May 2008. 95, 247

FITCH, W., HAUSER, M. & CHOMSKY, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition*, **97**, 179–210. 31

GALANTUCCI, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, **29**, 737–767. 42

GIL, D. (2008). How much grammar does it take to sail a boat?(or, what can material artifacts tell us about the evolution of language. In A.D. Smith, K. Smith & R. Ferrer i Cancho, eds., *The Evolution of Language. Proceedings of the 7th International Conference (EVOLANG 7)*, 123–130, World Scientific Press, Singapore. 79, 136

GILLIS, S. (1997). The acquisition of diminutives in Dutch. In W. Dressler, ed., *Studies in Pre- and Protomorphology*, 165–179, Verlag der Österreichischen Akademie der Wissenschaften, Vienna. 18

GILLIS, S., ed. (1998). *Studies in the Acquisition of Number and Diminutive Marking*, vol. 95 of *Antwerp Papers in Linguistics*. University of Antwerp, Antwerp. 18

GIVÓN, T. (1997). Grammatical relations: An introduction. In T. Givón, ed., *Grammatical Relations: A Functional Perspective*, vol. 35 of *Typological Studies in Language*, 1–84, John Benjamins, Amsterdam. 81, 87, 88, 89, 90

GIVÓN, T. (2001). *Syntax*, vol. 1. John Benjamins, Amsterdam. 59, 64, 268

GLEASON, H.A. (1965). *Linguistics and English Grammar*. Holt, Rinehart and Winston, New York. 84

GOLDBERG, A.E. (1995). *A Construction Grammar Approach to Argument Structure*. Chicago UP, Chicago. 28, 60, 69, 70, 71, 72, 73, 74, 75, 76, 96, 115, 123, 219, 247, 248, 249, 251, 256, 257

GOLDBERG, A.E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Science*, **7**, 219–224. 31

GOLDBERG, A.E. (2006). *Constructions At Work: The Nature of Generalization in Language*. Oxford University Press, Oxford. 71

GOLDBERG, A.E., CASENHISER, D.M. & SETHURAMAN, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, **15**, 289–316. 18, 71

GOULD, S. & VRBA, E. (1982). Exaptation. A missing term in the science of form. *Paleobiology*, **8**, 4–15. 31

GRUBER, J.S. (1976). *Lexical Structures in Syntax and Semantics*, vol. 25 of *North-Holland Linguistic Series*. North-Holland, Amsterdam. 58

HALLIDAY, M.A. (1994). *An Introduction to Functional Grammar*. John Benjamins, Amsterdam. 13, 16

HASPELMATH, M. (1998). Does grammaticalization need reanalysis? *Studies in Language*, **20**, 315–350. xix, 22, 135, 183, 185, 270, 271, 273, 275

HASPELMATH, M. (2003). The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In M. Tomasello, ed., *The New Psychology of Language*, vol. 2, 211–242, Lawrence Erlbaum, Mahwah, New Jersey. xviii, 87, 262, 263, 264, 265

HASPELMATH, M. (2007). Pre-established categories don't exist. *Linguistic Typology*, **11**, 119–132. 20, 60

HASPELMATH, M. (to appear). Terminology of case. In A. Malchukov & A. Spencer, eds., *Handbook of Case*, Oxford University Press, Oxford. 8

HASPELMATH, M., DRYER, M.S., GIL, D. & COMRIE, B., eds. (2005). *The World Atlas of Language Structures*. Oxford University Press, Oxford. 19

HAUSER, M.D., CHOMSKY, N. & FITCH, W.T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, **298**, 1569–1579. 31

HAWKIN, S. (1988). *A Brief History of Time. From the Big Bang to Black Holes*. Bantam Books, London, introduction by Carl Sagan. 46

HOEFLER, S. & SMITH, A.D. (2008). Reanalysis vs metaphor: What grammaticalisation can tell us about language evolution. In A.D. Smith, K. Smith & R. Ferrer i Cancho, eds., *The Evolution of Language. Proceedings of the 7th International Conference (EVOLANG 7)*, 34–41, World Scientific Press, Singapore. 126, 135

HOPPER, P. (1987). Emergent grammar. *BLC*, **13**, 139–157. 29

HOPPER, P. & TRAUGOTT, E. (1993). *Grammaticalization*. Cambridge University Press, Cambridge. 21, 182, 183, 185, 270

JACKENDOFF, R. (1990). *Semantic Structures*. MIT Press, Cambridge, MA. 63, 64, 268

JACKENDOFF, R. (1997). *The Architecture of the Language Faculty*, vol. 28 of *Linguistic Inquiry Monograph*. MIT Press, Cambridge MA. 13, 16, 17

JACKENDOFF, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press, Oxford. 41

JACKENDOFF, R. & PINKER, S. (2005). The nature of the language faculty and its implications for the evolution of language (reply to Fitch, Hauser, & Chomsky). *Cognition*, **97**, 211–225. 31

KAPLAN, F., MCINTYRE, A. & STEELS, L. (1998). An architecture for evolving robust shared communication systems in noisy environments. In *Proceedings of Sony Research Forum 1998*. 52

KAY, P. (2005). Argument structure constructions and the argument-adjunt distinction. In M. Fried & H.C. Boas, eds., *Grammatical Constructions: Back to the Roots*, 71–98, John Benjamins, Amsterdam. 247, 248

KAY, P. & FILLMORE, C.J. (1999). Grammatical constructions and linguistic generalizations: The what's x doing y? construction. *Language*, **75**, 1–33. 27, 28, 76, 95, 247, 249, 256

KEENAN, E. & COMRIE, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, **8**, 63–99. 248, 268

KEULEERS, E. & DAELEMANS, W. (2007). Memory-based learning models of inflectional morphology: a methodological case study. *Lingue e Linguaggio*, **6**. 28

KING, R.D. (1969). *Historical Linguistics and Generative Grammar*. Prentice Hall, New York. 21

KIRBY, S. (1999a). Learning, bottlenecks and the evolution of recursive syntax. In T. Briscoe, ed., *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, 173–204, Cambridge University Press, Cambridge. 236

KIRBY, S. (1999b). Syntax out of learning: The cultural evolution of structured communication in a population of induction algorithms. In D. Floreano, J. Nicoud & F. Mondada, eds., *Advances in Artificial Life: Proceedings of the 5th European Conference on Artificial Life*, 694–703, Springer, Berlin. 238

KIRBY, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, J. Hurford & M. Studdert-Kennedy, eds., *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, 303–323, Cambridge University Press, Cambridge. 39, 42, 200

KIRBY, S. (2001). Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, **5**, 102–110. 39, 78

KIRBY, S. & HURFORD, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi, eds., *Simulating the Evolution of Language*, 121–148, Springer Verlag, London. 39, 40

KIRBY, S., SMITH, K. & BRIGHTON, H. (2004). From UG to universals: Linguistic adaptation through iterated learning. *Studies in Language*, **28**, 587–607. 39

KUNO, S. (1973). *The Structure of the Japanese Language*. MIT Press, Cambridge, MA. 32

LAKOFF, G. (1974). Dialogue with H. Parret. In H. Parret, ed., *Discussing Language*, 151–178, Mouton, the Hague. 14

LAKOFF, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. The University of Chicago Press, Chicago. 13, 16, 257

LAMBRECHT, K. (1994). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representation of Discourse Referents*. Cambridge Studies in Linguistics 71, Cambridge University Press, Cambridge. 16

LANDAU, B. & JACKENDOFF, R. (1993). 'What' and 'where' in spatial language and spatial cognition. *Behavioural and Brain Sciences*, **16**, 217–238. 85

LANGACKER, R.W. (1977). Syntactic reanalysis. In C. Li, ed., *Mechanisms of Syntactic Change*, 59–139, University of Texas Press, Austin. 183, 270

LANGACKER, R.W. (1987). *Foundations of Cognitive Grammar. Volume 1*. Stanford University Press, Stanford. 13, 16, 28, 256

LANGACKER, R.W. (1991a). *Concept, Image, and Symbol: the Cognitive Basis of Grammar*. Cognitive Linguistics Research 1, Mouton de Gruyter, Berlin. 15

LANGACKER, R.W. (1991b). *Foundations of Cognitive Grammar. Volume 2: Descriptive Application*. Stanford University Press, Stanford. 70

LANGACKER, R.W. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer, eds., *Usage-Based Models of Language*, 1–63, Chicago University Press, Chicago. 28, 139, 180

LEVIN, B. & RAPPAPORT HOVAV, M. (1995). *Unaccusativity: At the Syntax-Lexical Semantics Interface*. MIT Press, Cambridge, MA. 59

LEVIN, B. & RAPPAPORT HOVAV, M. (1999). Two structures for compositionally derived events. In *Proceedings of SALT 9*, 127–144, Cornell Linguistics Circle Publications, Ithaca, NY. 100

LEVIN, B. & RAPPAPORT HOVAV, M. (2005). *Argument Realization*. Research Surveys in Linguistics, Cambridge University Press, Cambridge. 26, 58, 59, 60, 63, 69, 253

LEVINSON, S.C. (2003). *Space in Language and Cognition*. Language, Culture and Cognition 5, Cambridge University Press, Cambridge. 85

LEVINSON, S.C. & WILKINS, D., eds. (2006). *Grammars of Space. Explorations in Cognitive Diversity*. Language, Culture and Cognition, Cambridge University Press, Cambridge. 85

LI, C. & THOMPSON, S. (1976). Subject and topic: A new typology of language. In C. Li, ed., *Subject and Topic*, 458–489, Academic Press, New York. 79

LOETZSCH, M., VAN TRIJP, R., & STEELS, L. (2008). Typological and computational investigations of spatial perspective. In I. Wachsmuth & G. Knoblich, eds., *Modeling Communication with Robots and Virtual Humans*, LNCS 4930, 125–142, Springer, Berlin. 32, 52, 85, 238

MICHAELIS, L.A. (2003). Headless constructions and coercion by construction. In E. Francis & L. Michaelis, eds., *Mismatch: Form-Function Incongruity and the Architecture of Grammar*, 259–310, CSLI Publications, Stanford. 70, 71

MICHAELIS, L.A. (to appear). Complementation by construction. In M. Hauser, ed., *Proceedings of the Thirty-Second Annual Meeting of the Berkeley Linguistics Society*, BLS, Berkeley. 247, 251

MITCHELL, T.M. (1997). *Machine Learning*. McGraw-Hill Series in Computer Science, WCB/McGraw-Hill. 164

MITHUN, M. (2005). Beyond the core: Typological variation in the identification of participants. *International Journal of Linguistics*, **71**, 445–472. 269

MOSHINSKY, J. (1974). *A Grammar of Southeastern Pomo*. University of California Press, Los Angeles. 86

MOY, J. (2006). *Word Order and Case in Models of Simulated Language Evolution*. Ph.D. thesis, University of York. xix, 42, 200, 236, 237, 238, 239, 241, 243, 244, 245, 246

NEWMEYER, F.J. (1994). A note on Chomsky on form and function. *Journal of Linguistics*, **30**, 245–251. 31

NIVRE, J. (2006). *Inductive Dependency Parsing*. Springer, Berlin. 114

NOWAK, M.A. & KRAKAUER, D.C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences of the United Stated of America*, **96**, 8028–8033. 36, 37, 38, 42, 48, 76, 78, 200

NOWAK, M.A., PLOTKIN, J.B. & KRAKAUER, D.C. (1999). The evolutionary language game. *Journal of Theoretical Biology*, **200**, 147–162. 36, 37

NOWAK, M.A., KOMAROVA, N.L. & NIYOGI, P. (2001). Evolution of Universal Grammar. *Science*, **291**, 114–118. 36

NURNBERG, G., SAG, I.A. & WASOW, T. (1994). Idioms. *Language*, **70**, 491–538. 27

NUYTS, J. (1992). *Aspects of a Cognitive-Pragmatic Theory of Language. On Cognition, Functionalism, and Grammar*. John Benjamins, Amsterdam. 14, 26, 64

OUDEYER, P.Y. (2005). The self-organization of speech sounds. *Journal of Theoretical Biology*, **233**, 435–449. 47, 55

PALMER, F. (1994). *Grammatical Roles and Relations*. Cambridge UP, Cambridge. 59, 63, 79, 85

PIATTELLI-PALMERINI, M., ed. (1980). *Language and Learning. The Debate between Jean Piaget and Noam Chomsky*. Routledge, London. 17

PINKER, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge UP, Cambridge. 16, 63, 69, 78, 253

PINKER, S. (2003). Language as an adaptation to the cognitive niche. In M. Christiansen & S. Kirby, eds., *Language Evolution: States of the Art*, 16–37, Oxford University Press, New york. 30, 32, 36

PINKER, S. (2005). Clarifying the logical problem of language acquisition. *Journal of Child Language*, **31**, 949–953. 19

PINKER, S. (2007). *The Stuff of Thought. Language as a Window into Human Nature*. Penguin Books, London. 97

PINKER, S. & BLOOM, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, **13**, 707–784. 30, 31, 36

PINKER, S. & JACKENDOFF, R. (2005). The faculty of language: What's special about it? *Cognition*, **95**, 201–236. 31, 36

POLLARD, C. & SAG, I.A. (1994). *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago. 95, 102, 247, 249

ROBINS, R. (1967). *A Short History of Linguistics*. Longman, London. 16

SAG, I.A. (unpublished). Sign-based construction grammar. an informal synopsis. <http://lingo.stanford.edu/sag/papers/theo-syno.pdf>, Last accessed on 1 May 2008. 247, 251

SISKIND, J.M. (2000). Visual event classification via force dynamics. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 149–155. 131

SKOUSEN, R. (1989). *Analogical Modeling of Language*. Kluwer, Dordrecht. 28, 219, 220, 260

SMITH, A.D. (2003a). Intelligent meaning creation in a clumpy world helps communication. *Artificial Life*, **9**, 175–190. 39, 54, 145, 242

SMITH, K. (2003b). *The Transmission of Language: Models of Biological and Cultural Evolution*. Ph.D. thesis, University of Edinburgh, Edinburgh. 237, 239, 244

SMITH, K. (2008). Is a holistic protolanguage a plausible precursor to language? a test case for modern evolutionary linguistics. *Interaction Studies*, **9**, 1–17. 41

SMITH, K. & HURFORD, J.R. (2003). Language evolution in populations: Extending the iterated learning model. In W. Banzhaf, J. Ziegler, P. Dittrich, J. Kim & T. Christaller, eds., *Advances in Artificial Life: Proceedings of the 7th Eurpopean Conferenceon Artificial Life*, 507–516, Springer, 2003. 240, 244, 246

SMITH, K., KIRBY, S. & BRIGHTON, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, **9**, 371–386. 39, 42, 76, 200

SPERBER, D. & WILSON, D. (1986). *Relevance: Communication and Cognition*. Harvard University Press, Cambridge, MA. 139

STEEDMAN, M. (2000). *The Syntactic Process*. MIT Press, Cambridge, MA. 95

STEELS, L. (1996a). Emergent adaptive lexicons. In P. Maes, M.J. Mataric, J.A. Meyer, J. Pollack & S.W. Wilson, eds., *From Animals to Animats 4: Proceedings of the Fourth International Conference On Simulation of Adaptive Behavior*, 562–567, The MIT Press, Cambridge Ma. 53, 79

STEELS, L. (1996b). Perceptually grounded meaning creation. In M. Tokoro, ed., *Proceedings of the International Conference on Multi-Agent Systems*, 338–344, The MIT Press. 52, 133

STEELS, L. (1996c). A self-organizing spatial vocabulary. *Artificial Life*, **2**, 319–332. 52, 79

STEELS, L. (1997a). Constructing and sharing perceptual distinctions. In M.W. van Someren & G. Widmer, eds., *Proceedings of the European Conference on Machine Learning*, 4–13, Springer, Prague, Czech Republic. 52, 133, 266

STEELS, L. (1997b). Language learning and language contact. In W. Daelemans, A. van den Bosh & T. Weijters, eds., *Workshop Notes of the ECML/MLnet Familiarization Workshop on Empirical Learning of Natural Language Processing Tasks*, ECML/MLnet, Prague, Czech Republic. 52

STEELS, L. (1997c). The origins of ontologies and communication conventions in multi-agent systems. *Journal of Agents and Multi-Agent Systems*, **1**, 169–194. 52

STEELS, L. (1997d). The origins of syntax in visually grounded robotic agents. In M.E. Pollack, ed., *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*, 1632–1641, Morgan Kaufmann, San Francisco, CA. 53

STEELS, L. (1997e). Self-organizing vocabularies. In C.G. Langton, ed., *Proceeding of Alife V*. 51, 52, 79

STEELS, L. (1998a). Structural coupling of cognitive memories through adaptive language games. In R. Pfeifer, B. Blumberg, J.A. Meyer & S.W. Wilson, eds., *From Animals to Animats 5: Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, 263–269, The MIT Press, Cambridge Ma. 52

STEELS, L. (1998b). Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation. In J.R. Hurford, M. Studdert-Kennedy & C. Knight, eds., *Approaches to the Evolution of Language: Social and Cognitive Bases*, 384–404, Edinburgh University Press, Edinburgh. 51

STEELS, L. (1999a). How language bootstraps cognition. In I. Wachsmuth & B. Jung, eds., *Proceedings der 4. Fachtagung der Gesellschaft für Kognitionswissenschaft*, 1–3, Bielefeld. 52

STEELS, L. (1999b). The spontaneous self-organization of an adaptive language. In S. Muggleton, ed., *Machine Intelligence 15*, 205–224, Oxford University Press, Oxford. 52

STEELS, L. (1999c). *The Talking Heads Experiment. Volume 1. Words and Meanings*. Special pre-edition for LABORATORIUM, Antwerpen. 52, 53

STEELS, L. (2000a). The emergence of grammar in communicating autonomous robotic agents. In W. Horn, ed., *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI)*, 764–769, IOS Press, Berlin, Germany. 44, 53, 55

STEELS, L. (2000b). Language as a complex adaptive system. In M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. Merelo & H.P. Schwefel, eds., *Parallel Problem Solving from Nature*, LNCS 1917, 17–28, Springer-Verlag, Berlin. 23, 43

STEELS, L. (2000c). The puzzle of language evolution. *Kognitionswissenschaft*, **8**, 153–150. 52

STEELS, L. (2001a). Language games for autonomous robots. *IEEE Intelligent Systems*, 16–22. 52

STEELS, L. (2001b). The role of language in learning grounded representations. In P. Cohen & T. Oates, eds., *Papers from the 2001 AAI Spring Symposium*, AAAI Press, Menlo Park, CA. 52

STEELS, L. (2002a). Grounding symbols through evolutionary language games. In A. Cangelosi & D. Parisi, eds., *Simulating the Evolution of Language*, 211–226, Springer Verlag, London. 52

STEELS, L. (2002b). Simulating the evolution of a grammar for case. In *Proceedings of the Evolution of Language Conference*, Harvard. 125, 127, 130, 148, 152, 158, 166, 180

STEELS, L. (2003a). Intelligence with representation. *Philosophical Transactions of the Royal Society A*, **361**, 2381–2395. 51

STEELS, L. (2003b). Language re-entrance and the 'inner voice'. *Journal of Consciousness Studies*, **10**, 173–185. 54, 145, 242

STEELS, L. (2004a). The architecture of flow. In M. Tokoro & L. Steels, eds., *A Learning Zone of One's Own*, 135–150, IOS Press, Amsterdam. 54

STEELS, L. (2004b). The autotelic principle. In I. Fumiya, R. Pfeifer, L. Steels & K. Kunyoshi, eds., *Embodied Artificial Intelligence*, vol. 3139 of *LNAI*, 231–242, Springer, Berlin. 54

STEELS, L. (2004c). Constructivist development of grounded construction grammars. In W. Daelemans, ed., *Proceedings 42nd Annual Meeting of the Association for Computational Linguistics*, 9–19, Barcelona. 54, 125, 127, 152, 158, 166, 180

STEELS, L. (2005a). The emergence and evolution of linguistic structure: From lexical to grammatical communication systems. *Connection Science*, **17**, 213–230. 53

STEELS, L. (2005b). What triggers the emergence of grammar? In *AISB'05: Proceedings of EELC'05*, 143–150, AISB, Hatfield. 44, 53, 99

STEELS, L. (2006). How to do experiments in artificial language evolution and why. In A. Cangelosi, A. Smith & K. Smith, eds., *Proceedings of the 6th International Conference on the Evolution of Language*, World Scientific Publishing, London. 45, 46, 49

STEELS, L. (2007). The recruitment theory of language origins. In C. Lyon, C. Nehaniv & A. Cangelosi, eds., *The Emergence of Communication and Language*, 129–151, Springer Verlag, Berlin. 31, 32, 54, 127

STEELS, L. & BAILLIE, J.C. (2003). Shared grounding of event descriptions by autonomous robots. *Robotics and Autonomous Systems*, **43**, 163–173. 130, 131

STEELS, L. & BELPAEME, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, **28**, 469–529. 52, 238, 267

STEELS, L. & BLEYS, J. (2005). Planning what to say: Second order semantics for Fluid Construction Grammars. In A.J. Bugarín Diz & J. Santos Reyes, eds., *Proceedings CAEPIA 2005, 11th conference of the Spanish association for Artififcial Intelligence*, Lecture Notes in Artificial Intelligence, Springer, Santiago de Compostela. 54, 55

STEELS, L. & BROOKS, R.A., eds. (1995). *The 'Artificial Life' Route to Artificial Intelligence: Building Embodied, Situated Agents*. Lawrence Erlbaum Ass, New Haven. 51

STEELS, L. & DE BEULE, J. (2006). Unify and merge in Fluid Construction Grammar. In P. Vogt, Y. Sugita, E. Tuci & C. Nehaniv, eds., *Symbol Grounding and Beyond.*, LNAI 4211, 197–223, Springer, Berlin. 54, 103

STEELS, L. & KAPLAN, F. (1998a). Spontaneous lexicon change. In *Proceedings COLING-ACL 1998*, 1243–1250, Morgan Kaufmann, San Francisco, CA. 52

STEELS, L. & KAPLAN, F. (1998b). Stochasticity as a source of innovation in language games. In C. Adami, R.K. Belew, H. Kitano & C.E. Taylor, eds., *Proceedings of the Sixth International Conference on Artificial Life*, MIT Press. 52, 134

STEELS, L. & KAPLAN, F. (1999a). Collective learning and semiotic dynamics. In D. Floreano, J.D. Nicoud & F. Mondada, eds., *Advances in Artificial Life. Proceedings of the Fifth European Conference, ECAL'99*, vol. 1674 of *Lecture Notes in Computer Science*, 679–688, Springer, Lausanne, Switzerland. 52

STEELS, L. & KAPLAN, F. (1999b). Situated grounded word semantics. In T. Dean, ed., *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99)*, 862–867, Morgan Kaufmann, Stockholm, Sweden. 52

STEELS, L. & KAPLAN, F. (2002). Bootstrapping grounded word semantics. In T. Briscoe, ed., *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, 53–73, Cambridge University Press, Cambridge. 52

STEELS, L. & LOETZSCH, M. (2008). Perspective alignment in spatial language. In K.R. Coventry, T. Tenbrink & J.A. Bateman, eds., *Spatial Language and Dialogue*, Oxford University Press. 52, 267

STEELS, L. & MCINTYRE, A. (1999). Spatially distributed naming games. *Advances in Complex Systems*, **1**, 301–323. 52, 246

STEELS, L. & VOGT, P. (1997). Grounding adaptive language games in robotic agents. In P. Husbands & I. Harvey, eds., *Proceedings of the 4th European Conference on Artificial Life*, 473–484, The MIT Press, Brighton, U.K. 52

STEELS, L. & WELLENS, P. (2006). How grammar emerges to dampen combinatorial search in parsing. In P. Vogt, Y. Sugita, E. Tuci & C. Nehaniv, eds., *Symbol Grounding and Beyond. Proceedings of the Third EELC*, LNAI 4211, 76–88, Springer-Verlag, Berlin. 44, 53, 99, 128, 200

STEELS, L. & WELLENS, P. (2007). Scaffolding language emergence using the autotelic principle. In *IEEE Symposium on Artificial Life 2007 (Alife'07)*, 325–332, Wiley – IEEE Press, Honolulu, HI. 54, 128

STEELS, L., KAPLAN, F., MCINTYRE, A. & VAN LOOVEREN, J. (2002). Crucial factors in the origins of word-meaning. In A. Wray, ed., *The Transition to Language*, Oxford University Press, Oxford, UK. 52

STEELS, L., DE BEULE, J. & NEUBAUER, N. (2005). Linking in Fluid Construction Grammars. In *Proceedings of the 17th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC '05)*, 11–18, Brussels, Belgium. 54, 99

STEELS, L., VAN TRIJP, R. & WELLENS, P. (2007). Multi-level selection in the emergence of language systematicity. In F. Almeida e Costa, L.M. Rocha, E. Costa & I. Harvey, eds., *Advances in Artificial Life (ECAL 2007)*, LNAI 4648, 421–434, Springer, Berlin. 15, 53, 209

STEFANOWITSCH, A. & GRIES, S.T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, **2**, 209–243. 252

TAJIMA, M. (1985). *The Syntactic Development of the Gerund in Middle English*. Nan'un-do, Tokyo. 272

TATTERSALL, I. (1998). *The Origin of the Human Capacity*. American Museum of Natural History, New York. 13

TIMBERLAKE, A. (1977). Reanalysis and actualization in syntactic change. In C. Li, ed., *Mechanisms of Syntactic Change*, 141–177, University of Texas Press, Austin. 270

TOMASELLO, M. (1995). Joint attention as social cognition. In C. Moore & P.J. Dunham, eds., *Joint Attention: Its Origins and Role in Development*, Lawrence Erlbaum Associates, Hillsdale, NJ. 129

TOMASELLO, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press, Harvard. 30

TOMASELLO, M. (2005). Beyond formalities: The case of language acquisition. *The Linguistic Review*, **22**, 183–197. 17, 18

TRAUGOTT, E.C. (1992). Syntax. In R.M. Hogg, ed., *The Cambridge History of the English Language. The Beginnings to 1066*, vol. 1, 168–289, Cambridge University Press, Cambridge. 274

VAN DEN BOSCH, A. (1999). Instance-family abstraction in memory-based language learning. In I. Bratko & S. Dzeroski, eds., *Machine Learning: Proceedings of the Sixteenth International Conference (ICML'99)*, 39–48. 29

VAN DEN BOSCH, A. (to appear). Careful abstraction over instance families in memory-based language learning. *Journal of Experimental and Theoretical Artificial Intelligence. Special Issue on Memory-Based Language Processing*. 29

VAN DEN BROECK, W. (2007). A constraint-based model of grounded compositional semantics. In L. Lopes, T. Belpaeme & S.J. Cowley, eds., *Language and Robots: Proceedings of the Symposium*, Universidade de Aveiro, Aveiro. 55

VAN DEN BROECK, W. (2008). Constraint-based compositional semantics. In A.D. Smith, K. Smith & R. Ferrer i Cancho, eds., *The Evolution of Language. Proceedings of the 7th International Conference (EVOLANG 7)*, 338–345, World Scientific Press, Singapore. 55

VAN HOECKE, W. (1996). The Latin dative. In W. Van Langendonck & W. Van Belle, eds., *The Dative. Volume 1. Descriptive Studies.*, 3–37, John Benjamins, Amsterdam. 87

VAN LOOVEREN, J. (2005). *Design and Performance of Pre-Grammatical Language Games*. Ph.D. thesis, Vrije Universiteit Brussel, Brussels. 53, 155

VAN TRIJP, R. (2008a). Argumentsstruktur in der Fluid Construction Grammar. In K. Fischer & A. Stefanowitsch, eds., *Konstruktionsgrammatik II: Von der Konstruktion zur Grammatik*, vol. 47 of *Stauffenburg Linguistik*, Stauffenburg Verlag, Tübingen, translated by Arne Zeschel. 4, 247

VAN TRIJP, R. (2008b). The emergence of semantic roles in Fluid Construction Grammar. In A.D. Smith, K. Smith & R. Ferrer i Cancho, eds., *The Evolution of Language. Proceedings of the 7th International Conference (EVOLANG 7)*, 346–353, World Scientific Press, Singapore. 54

VAN VALIN, R.D. (1999). Generalized semantic roles and the syntax-semantics interface. In F. Corblin, C. Dobrovie-Slobin & J.M. Marandin, eds., *Empirical Issues in Formal Syntax and Semantics 2*, 373–389, Thesus, The Hague. 61, 63

VAN VALIN, R.D. (2004). Semantic macroroles in Role and Reference grammar. In R. Kailuweit & M. Hummel, eds., *Semantische Rollen*, 62–82, Narr, Tübingen. 61, 62, 64, 78, 268

VERHAGEN, A. (2006). Syntax, recursion, productivity – a usage-based perspective on the evolution of grammar. In A. Lubotsky, J. Schaeken & J. Wiedenhof, eds., *Evidence and Counter-Evidence. Festschrift F. Kortlandt*, vol. 1 of *SSGL 32*, Rodopi, Amsterdam. 28

VOGT, P. (2000). *Lexicon Grounding on Mobile Robots*. Ph.D. thesis, Vrije Universiteit Brussel. 52

VOGT, P. (2007). Group size effects on the emergence of compositional structures in language. In F. Almeida e Costa, L.M. Rocha, E. Costa & I. Harvey, eds., *Advances in Artificial Life (ECAL 2007)*, LNAI 4648, 403–414, Springer, Berlin. 44

WEINREICH, U., LABOV, W. & HERZOG, M. (1968). Empirical foundations for a theory of language change. In W. Lehman & Y. Malkiel, eds., *Directions for Historical Linguistics: A Symposium*, 97–195, University of Texas, Austin. 23

WELLENS, P. (2008). Coping with combinatorial uncertainty in word learning: A flexible usage-based model. In A.D. Smith, K. Smith & R. Ferrer i Cancho, eds., *The Evolution of Language. Proceedings of the 7th International Conference (EVOLANG 7)*, 370–377, World Scientific Press, Singapore. 179, 267

WELLENS, P., LOETZSCH, M. & STEELS, L. (2008). Flexible word meaning in embodied agents. *Connection Science*, **20**, 173–191. 43, 53, 83, 135

WILSON, D. & SOBER, E. (1994). Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences*, **17**, 585–654. 202

WRAY, A. (1998). Protolanguage as a holistic system for social interaction. *Language & Communication*, **18**, 47–67. 135

WRAY, A., ed. (2002). *The Transition to Language*. Oxford University Press, Oxford, UK. 135