Vrije Universiteit Brussel
Faculteit Wetenschappen
Departement Informatica

Adaptive Strategies in the Emergence of Lexical Systems

Proefschrift voorgelegd tot het behalen van de graad van doctor in de wetenschappen aan de Vrije Universiteit Brussel te verdedigen door

Pieter WELLENS

Promotor: Luc Steels
Brussels, February 2012

Print: Silhouet, Maldegem

2012 Uitgeverij VUBPRESS Brussels University Press
VUBPRESS is an imprint of ASP nv (Academic and Scientific Publishers nv)
Ravensteingalerij 28
B-1000 Brussels
Tel. +32 (0)2 289 26 50
Fax +32 (0)2 289 26 59
E-mail: info@vubpress.be
www.vubpress.be

ISBN 978 90 5487 965 7
NUR 984 / 616
Legal deposit D/2012/11.161/037

Acknowledgements

Although there can only be a single author of a thesis, it is never truly the work of a single person. For one, ideas don't belong to one person, they are born and shaped in discussion with colleagues and friends. At the VUB AI-Lab I had the pleasure to work with some of the sharpest and unrelenting minds I know. How memorable are some of the fierce discussions during our meetings, workshops or post-lunch coffee breaks.

I had the fortune to start my PhD around the same time as Joris Bleys and we immediately became "PhD companions". All great memories I have about my PhD involve Joris in one way or another. Thanks for making life at the lab as fun as it was and for being a great friend. Joachim De Beule, the Kurt Rogiers of science, has shared time and time again his profound insights about semiotic dynamics (and occasionally politics). Insights that have influenced this thesis more deeply than he might think. Katrien Beuls, the rising star of the AI-Lab (and beyond!), is, beyond doubt, the most active person I know. Cycling, wall climbing, PhD writing, ballroom dancing, you name it, she does it all. The labs latest recruit, Kevin Stadler, has, with his eager mind, more than once forced me to rethink some of the things I took for granted. Did I almost forget Frederik Himpe? Of course I did not! This system administrator is the true computer wizard of the AI-Lab. Each time I was in need of long forgotten command-line wisdom, Frederik was there to work his magic.

The AI-Lab attracts talents from beyond the reaches of computer science. The way in which both Carl Jacobs and Lara Mennes have managed and in their ways strengthened the Lab amazes and inspires me still. Thanks for your support both as friends and as colleagues.

I had the privilege to work closely with the brilliant people of Sony CSL Paris. Both Remi van Trijp and Martin Loetzsch started around the same time as I did and it was a pleasure to work closely with both of them. Thinking back to those early days

brings a spontaneous smile to my face. Remi is one of those rare computational linguists that actually knows how to code like a real Lisp hacker. That is rare. Martin is a true programming virtuoso with an exceptional eye for aesthetics. Not any of the graphs in this thesis would look as they do without him.

I remember very well, Michael and Simon joining the team during an off-site meeting, and later Vanessa, Nancy and Katya as part of the ALEAR Project. What a memorable three years that were. A special thanks to Vanessa and Nancy for their hospitality at the "Vanancy", including that time when I was locked out of the Sony apartment by, what must have been, a were-pigeon.

If there is a single person without whom this thesis would not have been possible it is my supervisor Luc Steels. Luc is one of those genuinely inspiring scientists able to show how science can be great. Interactions with Luc never fail to lead to stimulating ideas, or sometimes even radically new perspectives. Thank you for providing such a free, yet stimulating and supportive environment to conduct this research.

I am indebted to the members of my PhD committee, Andrew Smith and Paul Vogt as external reviewers and Ann Nowé, Bart de Boer and Tom Lenaerts as internal reviewers. Their excellent and in-depth feedback has substantially improved the thesis in terms of clarity and impact.

A thesis is not only written at work. You carry it with you most of the time, often working through the night or weekends at home. I want to thank Elfie Basselé, for all the support and love. And yes, also for spotting that persistent bug! My parents have always given me the freedom to pursue my own interests and my brother and sister have each influenced me in their own ways, more than they might think. So a big thanks to all of you. And although they might be small, my little nieces and nephew deserve their own place in these acknowledgements. Their smiles and spontaneous dances have cheered me up more than I can remember during these last few years.

Many breakthroughs only happen after taking a step back and not thinking about code and graphs and writing. So thankfully my dear friends, all the way in the far reaches of Antwerp, took up the task of luring me away from work and out of my apartment for drinks and laughs. Thanks Timothy, Dries, Steven, Ian, Stefan, Koen, Marie, Carl, Joris, Wim and others that I am forgetting.

The research reported in this thesis has been financially supported by FWOAL 610 “Digging for the roots of understanding” (DRUST) part of ESF Euroscores program EuroUnderstanding, FWOAL 328 “The evolution of grounded grammars and ontologies”, GOA3 “Origins of Language and meaning” and “EU301/EU FP7 Artificial Language Evolution on Autonomous Robots”(ALEAR). Pieter Wellens has also received funding from the VUB as a teaching assistant.

Antwerp, 13 February 2012

Abstract

This thesis can be situated in a new scientific paradigm that investigates the emergence and evolution of linguistic conventions through evolutionary language games. My contributions are mainly related to lexical conventions and the gradual development of cognitive structures.

Two approaches to lexical acquisition exist in the literature. The first posits a strong division between concept formation and lexical learning. Concepts arise independently of language and the lexical acquisition problem can be reduced to a problem of mapping words to concepts. In the third chapter I present an overview of the cross-situational learning (XSL) approaches that have been successfully applied to this problem in the past.

The second view of lexical learning sees lexical acquisition as a process of creating and gradually shaping the meanings of words. In the fourth chapter I show that traditional XSL is not well suited for this more difficult task and that this second view of word learning needs more powerful processing and learning strategies.

I argue that, surprisingly, we should look at theories of grammar, such as cognitive grammar and grammaticalization theory, for inspiration. In chapter five I propose a strategy that is in line with the basic tenets of these theories and which addresses three shortcomings of the traditional XSL approach. First it no longer relies on an explicit enumeration of competing hypotheses, only a single hypothesis per word is maintained at all time. Second, the manner in which word meanings are used in processing stresses flexible re-use. Third, word meanings are internally adapted based on the feedback from flexible processing. I show that strategies embodying these tenets cope much more naturally with the problem of lexical acquisition and conventionalization when meanings have not been established beforehand.

The final part of the thesis ventures beyond the lexical domain toward grammatical categorization. The main research question adressed is whether the adaptive lexical learning algorithms proposed in the earlier chapters can be extended to also deal with the problem of semantic grammatical categorization and the new conventionalization problems this task brings along.

An overarching contribution, which motivated primarily the earlier chapters of the thesis, is to offer an overview of non-grounded lexical language games and show the progress that has been made over the past fifteen years. A substantial selection of strategies has been reimplemented so that, for

the first time, the strategies are explained and compared within the same multi-agent setting, using the same data and the same measures.

Samenvatting

Deze thesis kan gesitueerd worden binnen een nieuw wetenschappelijk paradigma dat het ontstaan en de evolutie van talige conventies onderzoekt via evolutionaire taalspelen (*language games*). Mijn bijdragen hebben voornamelijk betrekking op lexicale conventies en de stapsgewijze ontwikkeling van cognitieve structuren.

In deze thesis maak ik een onderscheid tussen twee benaderingen tot het probleem van lexicale taalontwikkeling. De eerste benadering poneert een strict onderscheid tussen conceptverwerving en lexicale taalverwerving. Concepten ontstaan grotendeels onafhankelijk van taal en als dusdanig kan het taalverwervingsprobleem gereduceerd worden tot het verbinden van woorden met concepten. In het derde hoofdstuk geef ik een overzicht van intersituationele leeralgoritmen (*cross-situational learning*) die in het verleden reeds met succes toegepast werden op dit probleem.

De tweede benadering tot lexicale taalverwerving benadrukt een proces van constructie en graduele vorming van de betekenis van woorden. In het vierde hoofdstuk toon ik aan dat traditionele intersituationele leeralgoritmen niet geschikt zijn om deze benadering tot taalverwerving overtuigend te modeleren.

Ik argumenteer dat een oplossing tot dit lexicale probleem, eerder verrassend, kan gevonden worden in grammaticale theoriën, zoals cognitieve grammatica en grammaticalisatie. In het vijfde hoofdstuk stel ik een strategie voor die compatibel is met de centrale principes van deze theoriën en die drie tekortkomingen van de traditionele aanpak aankaart. Ten eerste maakt deze strategie geen oplijsting van concurrerende hypothesen. Een woord kan slechts geassocieerd zijn met één betekenis. Ten tweede benadrukt deze strategie flexibel hergebruik van reeds bestaande woorden en ten derde worden betekenissen intern aangepast, gebaseerd op dit flexibel hergebruik. Ik toon aan dat dergelijke strategiën op een meer natuurlijke wijze omgaan met het probleem van lexicale taalverwerving en conventionalizatie.

In het laatste deel van de thesis begeef ik me buiten het lexicale domein in de richting van grammaticale categorisatie. De onderzoeksvraag is of de adaptieve leeralgoritmen uitgebreid kunnen worden naar het domain van grammatica en of ze het hoofd kunnen bieden aan nieuwe conventionalisatieproblemen die de stap naar grammatica met zich meebrengt.

Een overkoepelende bijdrage, die voornamelijk tot uiting komt in de eerste hoofdstukken, bestaat uit een overzicht van lexicale taalspelen. Dit overzicht toont de vooruitgang geboekt gedurende de afgelopen vijftien jaar. Hiertoe

heb ik een uitgebreide selectie aan strategiën opnieuw geïmplementeerd, wat me ook toeliet om de strategiën uit te leggen en te vergelijken in eenzelfde multi-agent setting, gebruik makende van dezelfde data en parameters.

Contents

List of Figures

List of Tables

Part I

Introduction

Chapter 1

Introduction

As in any scientific endeavour this thesis represents only a small part of a much larger effort. In this case the larger effort is understanding the dynamics underlying the conventionalization of linguistic phenomena. The research reported in this thesis is best situated in the field of *evolutionary linguistics* which posits that to understand language we need to understand its emergence and evolution.

Evolutionary linguistics cuts through many disciplines ranging from psychology, linguistics and anthropology to biology, physics and computer science. However diverse these backgrounds may be, evolutionary linguistics brings them together by supplying a family of concepts and ideas. In this shared view, language is regarded as a *complex adaptive system*, used for communication and thus rooted in interaction. In communication language serves a function, speakers of a language can *do* things with language, reach goals with it. Not only must our unique language capacity be explained in terms of genetic biological evolution but equally well in terms of *cultural* evolution (Steels, 2004b; Brighton *et al.*, 2005; Kirby *et al.*, 2008). Evolutionary linguistics also subscribes to the *usage-based* approaches to language. Actual situated language use is both crucial in learning but also in change, it is the driving force of the conventionalization processes that have shaped and continue to shape today's languages.

This thesis serves two goals. It presents an *overview* of the most important lexical language games that have been investigated within the language game paradigm over the past fifteen years. For each of these games different strategies are presented and contrasted. All these language games and strategies have been reimplemented so that they can be explained and compared within the same multi-agent setting and using the same data. Secondly this thesis also introduces original *new strategies*, most prominently within the domain of multi-word utterances and multi-dimensional meanings. I propose an elegant

and scalable solution which relies on flexibility in processing and adaptivity in representation. It stresses the need for reuse of existing means of expression and shows how this reuse impacts the language itself.

In this introductory chapter we introduce the main ideas of evolutionary linguistics and the methodology followed in this thesis.

1.1 Language as a complex adaptive system

Language is often studied as a static system of words and rules (Pinker, 1998; Chomsky, 1965). A static view can only capture a snapshot of language and it turns out many answers to the puzzles of language are to be found in its dynamic and ever changing nature. Evolutionary linguistics does not share the static view, but instead views language as a *complex adaptive system*.

A complex system is a system from which the behaviour cannot be trivially deduced from the rules followed by its constituent parts. In this respect language is indeed a *complex* system because it is the product of conventionalization processes over long periods of time, involving often large populations of language users. New manners of expression come and go and their conventionalization is often the byproduct of situated language use.

Language is not just a complex system, it is a complex *adaptive* system (Steels, 2000; Briscoe, 1998; Beckner *et al.*, 2009) because it changes in response to the communicative needs of its speakers. This adaptivity is at play on different interconnected levels. At the level of the conventionalized language of the population as a whole, change might not be so apparent. Nevertheless languages gradually change over time, bringing with it the emergence of new forms of expression and the disappearance of others (McMahon, 1994).

At the level of an individual agent the change is more apparent. The linguistic inventory of an agent is always in flux, new words are picked up, metaphors are created and stretched. Most of those innovations never even reach conventional status at the population level.

These entwined aspects make it enormously difficult to predict the impact of a cognitive capability in the long term conventionalization process in the entire population. Especially when language is viewed as a complex adaptive system computational modeling becomes a crucial tool in understanding the dynamics at play.

1.2 The semiotic cycle

Evolutionary linguistics posits a strong bond between language and communication. Language is used to communicate, to interact, to reach goals. An

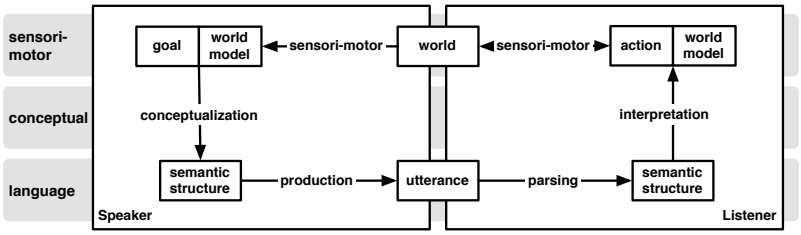


Figure 1.1: Semiotic cycle representing the relationship between interlocutors in a routinized interaction.

interaction takes place between at least two interlocutors and is situated in a context. It involves both the production of utterances and their interpretation and understanding. All these aspects are important in the study of language and for understanding its emergence and evolution.

The *semiotic cycle* captures the relationship between speaker, listener and the outside world as an abstraction underlying basic forms of interaction (see Figure 1.1). Two interlocutors, a speaker and a listener, pass through three levels of processing: sensori-motor, conceptual and linguistic. Both participants build a world model using their sensors, previous experiences, ontologies, etc. It is assumed that language serves a function, the speaker wishes to reach a goal and uses language as a means to reach it. Given his world model and a goal the speaker *conceptualizes* a semantic structure which serves as input to the linguistic system *producing* a corresponding utterance.

In all but the simplest interactions there can be a multitude of different conceptualizations for reaching the same goal. Likewise a single semantic structure can be expressed in different ways, some more conventional than others or some with a higher chance of reaching the projected goal. For example, you are in a restaurant and would like some more water. Depending on the situation this can be expressed in different ways:

- Could someone please pass me the bottle of water.
- Peter, the water please.
- (pointing) yeah, the water, yeah. Thanks.
- (after eating a hot pepper) Water. Now!

Such differences in conceptualization and production have been famously investigated using the pear stories (Chafe, 1980). Different audiences were asked to describe the same seven-minute film which showed a narrative

involving a man picking pears and events surrounding this. It was shown that cultural, contextual and cognitive factors all influenced the way the subjects retold the story (Tannen, 1980; Downing, 1980; Clancy, 1980).

The semiotic cycle continues with the listener *parsing* the utterance to a semantic structure. During parsing the listener can take parts of the current situation into account but it is only in *interpretation* that the semantic structure gets fully grounded in the current context. The listener can then decide to act upon the utterance which in turn allows the speaker to determine whether his goal has been reached.

The semiotic cycle as such remains abstract and schematic. It is operationalised and further extended by concrete *language games*, which are small scripted interactions.

1.3 Language games and conventionalization problems

Following Wittgenstein (1967) we take a simplified interaction as one of the main tools to investigate linguistic behaviour. These interactions (involving at least one listener and one speaker) are called *language games*. Language games help to maintain a balance between simplicity in modeling and complexity of situated language use and change. Every language game, no matter how stripped of its resemblance to a real world setting, subscribes to the importance of *interaction* for understanding language. In these interactions agents wish to reach a goal, making the interaction purposeful and meaningful. Most language games also stress the importance of situatedness in a shared context and of joint attention (Moore & Dunham, 1995; Diessel, 2006). Often, but not necessarily, agents are embodied in the real world through the use of robots further widening the explanatory potential of the paradigm.

Although most particular language games have a prototypical interaction script associated with them, I argue in this thesis that the interaction script alone is not the best way to distinguish or classify different language games. I believe that the type of *conventionalization problem* that the agents need to solve is a better candidate for classification and should be the primary motivation to differentiate between language games. In this thesis I introduce different conventionalization problems and associate a language game with each one of those. For example for Naming Games as defined in Chapter 2 the only conventionalization problem is that of damping different names for the same object, a problem I call *word from competition*. The problem of word form competition can however be evoked by different interaction scripts. Nevertheless according to my classification it is the conventionalization

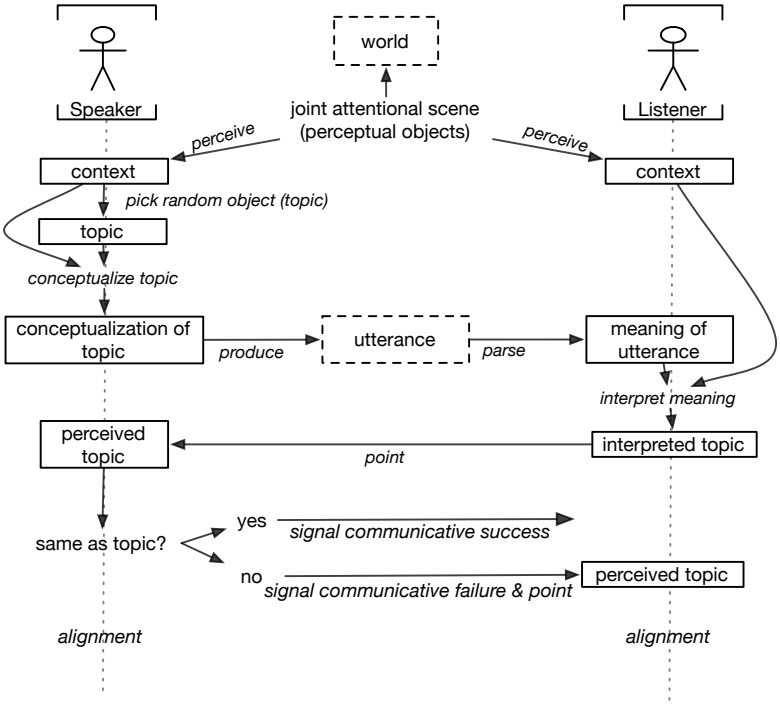


Figure 1.2: A schematic representation of the flow of a typical language game. Not every type of language game operationalizes every step in this picture. For example, a Naming Game does not require conceptualization.

problem that determines whether a language games is a Naming Game or not and not its interaction script.

This is not to say the more well-known language games have a well-established script that is followed by most implementations. For example a language game usually has only two participants, a speaker and a listener. The typical script of the games discussed in this thesis consists of a speaker trying to draw the attention of a listener to an object or an ongoing event in the shared context by describing it in language. The agents, either simulated or embodied, have extralinguistic means of communication as well, such as nodding and pointing. These feedback mechanisms allow for learning and alignment. A schematic representation is shown in Figure 1.2.

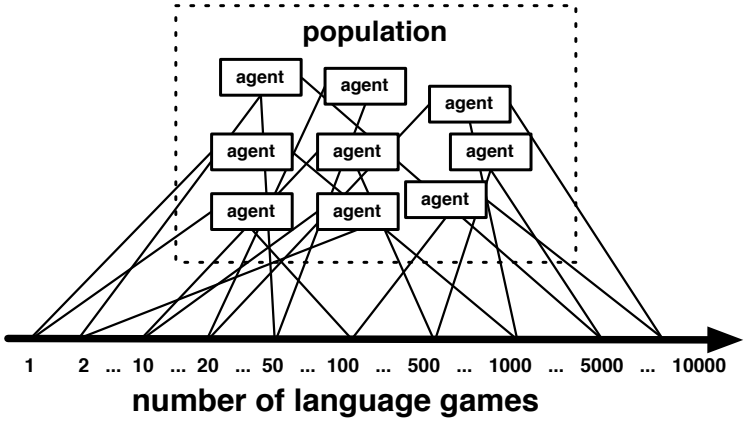


Figure 1.3: An experiment consists of a series of thousands of language games. Every game is played by two randomly picked agents from the population. Each game is local to only the two participating agents.

The language game paradigm (Steels, 1995; Nowak *et al.*, 1999; Steels, 2002) presupposes a multi-agent setup (Ferber, 1999) with at least, but generally more than, two agents. In a typical experiment the agents start without any shared knowledge of the world or of language. In fact there is no language at all, the agents have to develop a shared language system by playing language games. There is no leader or teacher so the agents can only achieve this by engaging in communicative interactions. For all reported experiments we use the multi-agent framework, Babel 2 (Loetzsch *et al.*, 2008b).

Information in a language game is *local* to its participants, which means the other agents are not aware of innovations or adaptations that might have come up during that interaction. Only after the innovation is used again in other interactions can it spread through the population. In a language game experiment a population of agents plays thousands of language games 1.3.

1.3.1 Routine and meta layer processing

In the language game paradigm processing can take place on two distinct layers, the routine layer and the meta layer (Beuls *et al.*, 2012b; van Trijp, 2012; Steels & van Trijp, 2011). The *routine layer* corresponds to normal, unproblematic processing. The language game script shown in Figure 1.2

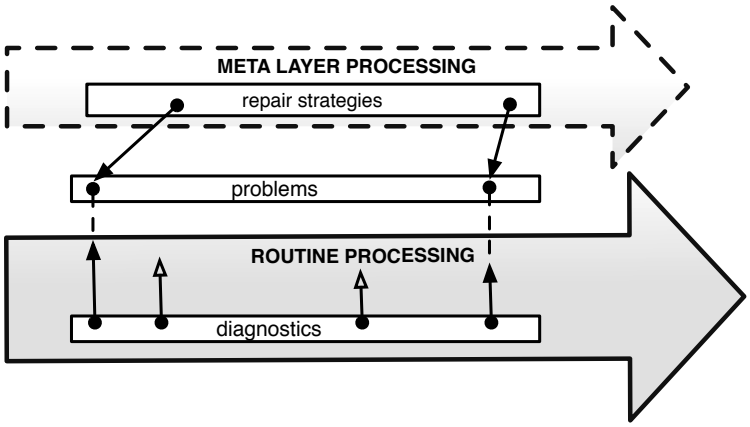


Figure 1.4: Schematic overview of routine and meta-layer processing in the language game paradigm. Diagnostics run during routine processing, potentially reporting problems in the process. Repair strategies, running in the meta-layer, can act on these problems and try to resolve them.

only shows this routine processing. Learning, which comprises both invention of new words and adoption of heard words, takes place at a second layer of processing called the *meta layer*, which is the layer agents resort to when encountering problems during processing.

During any step of the routinized script, problems can be encountered. For example the speaker's lexicon might not be expressive enough to describe the conceptualized topic, requiring him to extend his lexicon. Another example, this time for the listener, is the first exposure to a novel word. In such cases it is clear that routine processing cannot continue and agents need to resort to a more creative, problem-solving way of processing.

Different studies in neuroimaging show that additional brain activity in regions not previously active can be measured when subjects are exposed to ambiguous or incomplete utterances (Zempleni *et al.*, 2007). Ketteler *et al.* (2008) even assert that "Subcortical neural circuits probably become activated when the language processing system cannot rely entirely on automatic mechanisms but has to recruit controlled processes as well". These findings give credence that in non routine situations different mechanisms come into play.

Agents thus need a capacity to detect problems in processing, called diagnostics, and to repair or circumvent them in meta-layer processing, called

repair strategies.

1.3.2 Diagnostics and re-entrance

The task of detecting problems during routine processing is the responsibility of *diagnostics*. Diagnostics are small processing entities that are active during routine processing and, at regular intervals, diagnose for a specific problem. The diagnostic either reports that all is well, in which case routine processing continues, or it reports a specific problem. When a problem is reported, routine processing also continues, with the difference that the agent is now aware that a problem exists.

Diagnosed problems do not repair themselves, so a final type of processing entity, a *repair strategy* is needed. Repair strategies operate on reported problems and, as their name suggests, try to repair the problem. For example repair strategies can try to extend the linguistic inventory of the agent, or they could change the manner of processing to allow more creative language use. Repairing is part of the meta layer and is thus not part of routine processing. Furthermore, repair strategies are not required to immediately act on a problem since they might require extra information in order to come up with a suitable solution. For example a listener confronted with a novel form of expression might diagnose this as problematic. The repair strategy acting on this problem is best to wait for non linguistic feedback, such as pointing, in order to limit the number of possible interpretations of the novel expression.

In a typical lexical game there are often only two diagnostics, one for the speaker and one for the listener. A speaker should be able to extend his lexicon by introducing new words when he is unable to express the conceptualised semantic structure. Likewise, as listener, an agent should be able to acquire new words. The speaker can detect a need for more expressive power by interpreting his produced utterance before actually uttering it. This allows him to verify whether his own interpretation leads him to the topic. As such the speaker takes himself as a model for the listener. This process of interpreting your own utterance is called *re-entrance* (Steels, 2003) and is schematically shown in Figure 1.5. Re-entrance rests on the same principles as the *Obverter* strategy introduced by Oliphant & Batali (1997) and later incorporated in the models of Kirby (2002). Failure of re-entrance triggers the speaker to extend his lexicon by inventing a new word meaning association.

The listener diagnoses for novel words in the utterance. If one or more unknown words are encountered in the utterance a problem is reported. Although the problem is easily diagnosed, finding a suitable meaning for the unknown words is the crux of a lexical language game.

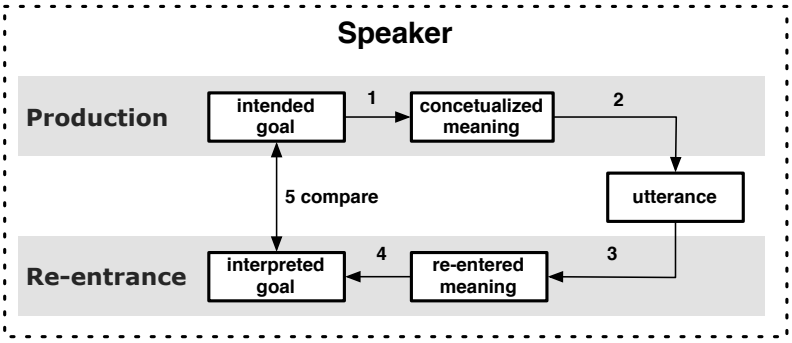


Figure 1.5: Schematic overview of re-entrance. All five steps are internal to the speaker, starting with normal production involving conceptualization (step 1) and verbalization (step 2). The utterance is then parsed by the speaker (step 3) which leads to a meaning and finally, after interpretation, to an interpreted topic (step 4). In diagnosing the speaker then compares his initially chosen topic to the interpreted topic (step 5). A problem is reported if they are different.

1.3.3 Alignment

With diagnostics and repair strategies agents can extend their inventory of constructions and conceptualizations by processes of invention and adoption. These processes alone, however, will not make a population of agents converge on a shared, coherent language system. For this we need a more subtle type of adaptation called *alignment*.

After playing thousands of language games, the agents in the population should converge on a common language system. As opposed to invention or adoption, which extend the linguistic inventory of the agent, alignment works on the constructions used during the language game. It takes place at the end of the language game after all other processing has finished. At that time the participating agents have all available information such as extra linguistic feedback or whether the game was a success.

Acting on this information, agents slightly modify their inventory with the aim of increasing or maintaining communicative success in future interactions. These modifications can range from adapting scores to removing forms of expression. Agents have only limited and indirect information about each other. Often in lexical language games the only input for alignment is whether the game was a success, which words were used, and which object was intended

by the speaker.

1.4 Language strategies and language systems

In order to render computational investigations into language evolution feasible we do not tackle all of language but instead focus on particular subsystems of language, called *language systems*. A language system focuses only on a small set of features of a mature language. The driving question is what minimal ingredients a language user needs in order to learn and use a particular subsystem. These ingredients are captured by a *language strategy* (Steels, 2012b; Bleys, 2010), a concept which is increasingly used to guide investigations into the evolution of language (see (Steels, 2012a) for a recent overview).

A wide variety of language systems has already been investigated and was recently bundled in (Steels, 2012a). For example Steels & Loetzsch (2012) investigates how a system of names for objects can emerge in a embodied setting. Bleys (2012) investigates different language strategies that could explain the emergence of a language system for describing colors. The subsystem of color terms has received maybe the most attention of all language systems in the language game paradigm (Belpaeme, 2002; Steels & Belpaeme, 2005; Belpaeme & Bleys, 2005). Another system that has received considerable attention is that of spatial language for describing objects in space (Steels & Loetzsch, 2008; Spranger *et al.*, 2010; Spranger, 2012).

Obviously these language systems can be primarily grammatical in nature as well. Case systems for marking event structure have been investigated in depth through computational modeling (see (van Trijp, 2011; van Trijp, 2008a,b)) and more recently also agreement (Beuls *et al.*, 2012a), aspect systems (Gerasymova *et al.*, 2009) and systems for quantification (Pauw & Hilferty, 2012) have been investigated in this methodological paradigm.

In each of these investigations one of the driving questions is what kinds of strategies can explain the emergence of the linguistic subsystems under investigation. The strategy itself has to describe the (minimal) representational capacities and all the necessary processing components.

The representational aspects of a language strategy determine how the agents represent form and semantics, how these are associated in constructions and potentially even in what type of structure the constructions are maintained (e.g. a set, a network, ...). For example for a grounded Naming System, robots need to be able to establish object identity over time. This requires that an agent can store multiple views of the same object, leading to a prototypical view of an object which is finally associated with a particular name (Steels & Loetzsch, 2012). van Trijp (2011) suggests that agents need

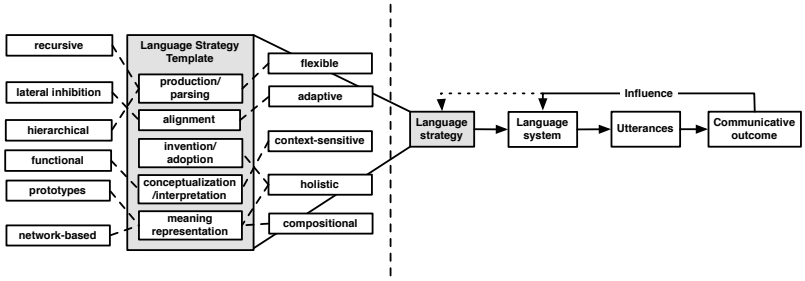


Figure 1.6: The left-hand part shows a schematic representation of a language strategy (the language strategy template) with parameters for processes such as production, alignment and learning. These parameters can then be instantiated in many different ways by different actual strategies leading to a particular instantiated language strategy. What is shown here is only an example, not the depiction of a particular strategy. Shown at the right (taken from (Steels, 2012b)) is how a language strategy fits in the larger picture of language use and change. A strategy thus gives rise to a particular language system, which leads to an associated verbal behaviour and communicative outcomes. These outcomes feed back into the language system and potentially in the long term even to the language strategy.

to be able to represent distinctive feature matrices in order to represent a mature case system, like the one found in German.

Another important part of representation is what kind of, if any at all, statistical scoring mechanism the agents require to keep track of certain usage-based properties of their linguistic inventory. For example a score representing the amount of times a construction was used in a successful language game. At which level of granularity and which types of scores are kept is also part of the representational aspects of the language strategy.

At the processing level a language strategy encompasses everything a single agent needs to successfully play a particular language game. Generally this encompasses production, interpretation, conceptualization, learning and alignment. These processes thus cut through the complete semiotic cycle (see Figure 1.1) and affect nearly every aspect of the language game. Furthermore, the processing components of a strategy influence each other. For example a more expressive language processing component for production lessens the need for invention of new linguistic entities. Equally more powerful repair strategies might require less powerful alignment strategies. Representational aspects of a strategy should also influence processing. For example a more

fine grained scoring strategy might lead to more subtle language processing and alignment. Figure 1.6 schematically shows a language strategy template and possible choices for instantiation of sub-components. The right-hand side of this figure shows how a template fits in the larger whole of language use and change.

In many cases different strategies can solve the same conventionalization problem.

1.5 Situating this research: linguistics, evolutionary biology and computer science

In trying to explain conventionalization processes in language this thesis brings together the fields of cognitive linguistics, evolutionary biology and computer science.

1.5.1 Cognitive Linguistics

For an outsider, linguistics is not an easy field to enter for there seems to exist little consensus on what linguistics is or should be. Linguists are divided into *generative* linguists, *cognitive* linguists, *minimalists*, *computational* linguists, each of those with an often radically different conception of language.

For example, summarizing what the generative tradition stands for is not as straightforward as one would hope because its founding father, Noam Chomsky, has revised the theory multiple times (Chomsky, 1965, 1980; Chomsky & Lasnik, 1993; Chomsky, 1995). Predominantly generative linguists search for universal features of language and are interested in describing the abstract processing system which is believed to underlie our actual performance. Instead of focusing on this real world performance they focus on *competence*, which is a speaker's perfect knowledge of the language. Generative linguist tend to pay more attention to syntax and morphology than to semantics and pragmatics.

The cognitive-functional approach searches for answers by widening the scope of investigation and altering the focal point from only syntax to also incorporate semantics, pragmatics, conceptualization and real (as opposed to perfect) situated language use. The most dominant approach to grammar emerging from this new school of thought is called *Construction Grammar*. Although different flavours of construction grammar exists, they generally agree on the following points (Goldberg, 2003; Croft & Cruse, 2004):

- All linguistic knowledge, both grammatical and lexical, is captured in *constructions*, which are pairings of form and function/meaning. Form

can range from morphemes and words to abstract phrasal patterns.

- These constructions are maintained in a network structure although the exact nature of the relationships is not fully agreed upon (Croft & Cruse, 2004). To share a language between language users is thus to share a conventionalized set of constructions.

- Learning or acquiring these constructions is driven, not by innate mechanisms, but by general cognitive mechanisms.The learning and construction of languages is driven by actual real world language use, leading to what is known as the *usage-based approach* to language learning (Tomasello, 2003; Barlow & Kemmer, 2000).

This thesis has been especially influenced by the work of Ronald Langacker, the founding father of cognitive grammar and cognitive linguistics in general (Langacker, 1987, 1991a, 2000a). In his view language is a dynamic system at every level of description. Langacker (2000a) stresses the importance of several psychological phenomena operationalized in later chapters of this thesis:

Entrenchment (also routinization, automization) refers to the observation that the occurrence of psychological events leaves a trace that facilitates later re-occurrence.

Abstraction is the process where a commonality is found among multiple experiences. The abstraction process thus "filters out" those facets of the individual experiences which do not recur.

Comparison of two structures consists of finding their commonalities and differences. In computer science this is often operationalized using a similarity measure. Comparison is a prerequisite for abstraction.

Although most reported work lies in the lexical domain another source of inspiration is *grammaticalization theory* (Heine *et al.*, 1991; Hopper & Traugott, 1993; Heine & Kuteva, 2007). This theory, supported by an ample amount of evidence, shows that new grammatical conventions emerge out of lexical constructions. The process of grammaticalization encompasses four interrelated processes, two semantic and two form-related (Heine & Kuteva, 2007). The two semantically motivated processes are *extension* and *desementaticization*. Extension is the process in which (conventionalized) linguistic expressions are extended to new (uncoventionalized) contexts. Desemanticization or *semantic bleaching* refers to a loss (or generalization) in meaning content and follows extension diachronically. The (repeated) act of extension provides the basis for the semantic bleaching.

The theory of grammaticalization is surprisingly compatible with the insights from cognitive grammar. For example an act of extension must to some extent rely on the capacity of comparison. Semantic bleaching requires both comparison and abstraction. Entrenchment also plays a crucial role in the theory of grammaticalization. When an extended use becomes used more often, entrenchment makes sure the novel way of expression becomes part of convention.

1.5.2 Evolutionary Biology

In the search for a theory of language evolution, evolutionary biology is bound to play an important role. What evolutionary change was responsible for our unique capacity of language? In what ways does our evolved biological constitution constrain or facilitate our linguistic capacities? This thesis, and in fact most of the language game methodology, does not answer these types of questions. The explanatory power of the language game paradigm resides within the domain of cultural evolution and much less in that of biological evolution. Languages differ radically from culture to culture and such a diversity cannot be explained by biological evolution alone (Evans & Levinson, 2009).

This thesis is only concerned with language as a cultural phenomenon undergoing cultural evolution and change. The impact of evolutionary biology comes more from the methodological framework and concepts it provides [see (Maynard-Smith & Szathmáry, 1995) for a survey]. Steels (1998b) convincingly argues that concepts like selection, level formation, self-organisation and co-evolution can be applied in the cognitive-cultural domain as well. In the same line Ferrer-I-Cancho & Forns (2010) investigates the parallel between self-organization in genomes and human languages.

1.5.3 Computer Science and Artificial Intelligence

Small changes in a language strategy might bring about massive, unexpected changes in the emerging language system. Given a sufficiently powerful model, computers can help us in exploring these issues. In the process we often discover that important factors were glossed over or the opposite, that aspects deemed crucial to the dynamics are in fact peripheral. Time and time again, during the course of this doctoral investigation, I have been amazed to see results of computational simulations contradict (my own) previously held intuitions.

Computational models have the added benefit that they require the author to be specific and unambiguous. They can be reimplemented, compared,

validated or falsified and they can support or contradict existing theories or models.

The language game paradigm is not the only computational paradigm investigating the evolution of language. For example the Iterated Learning Model [see (Smith *et al.*, 2003) for an overview] has been used to investigate the emergence of compositionality and structure in general (Kirby, 2001; Brighton, 2002; Vogt, 2005a; Kirby *et al.*, 2004). The iterated learning paradigm and the language game paradigm can be seen as complementary. Iterated learning tends to focus more on vertical, generational transmission of language whereas the in language games the focus is more on horizontal peer-to-peer transmission, although Scott-Phillips & Kirby (2010) stresses that vertical transmission is not definitional for iterated learning. Smith (2007) is an example that combines the two approaches.

1.5.4 Experimental Semiotics

More recently the cultural evolution of language is also being studied experimentally, often in combination with or inspired by computational modeling approaches (Scott-Philips & Kirby, 2010; Galantucci & Garrod, 2011; Galantucci, 2005; Garrod & Pickering, 2009; Theisen *et al.*, 2010; Cornish *et al.*, 2009). This has lead to the relatively new field of experimental semiotics [see (Galantucci & Garrod, 2011) for an overview] which focuses on the experimental investigation of novel forms of communication with human subjects.

The communication between computational language game modelling and experimental semiotics goes in two ways. Often, games that are first implemented computationally are later also investigated in the lab by experimental semioticians. On the other hand, the results coming from the field of experimental semiotics not only serve as inspiration but often also require computational modelers to rethink some of their assumptions. See (Galantucci & Steels, 2008) for an example of such a bi-directional interaction between the two fields.

1.6 Overview of this thesis: objectives and hypotheses

Due to the interdisciplinary nature of the thesis, different types of readers, with backgrounds ranging from linguistics and psychology to computer science, might read this thesis. I present in this section a short overview of the thesis and also point out which parts might be more interesting depending on the

background knowledge.

To frame the arguments I make it is important to realize that two different views of word learning exist in the literature. The dominant view depicts word learning as a task of mapping words to concepts. Meaning itself is not at the forefront of this view (Bloom, 2000). The other, more recent view, puts meaning at the center of the investigation and sees the gradual shaping and fine-tuning of meaning as the core problem of lexical learning (Bowerman & Choi, 2001; Tomasello, 2001).

As the title of the thesis suggests my original contributions are related to the adaptive strategies which are covered in chapters five to eight. Chapters two to four mainly focus on previous models but are are an integral part of the complete story I wish to tell. If you are only interested in adaptive strategies you should skip to Chapter 5, although some backtracking is bound to be necessary. For example the Compositional Guessing Game and its main conventionalization problem is introduced in Chapter 4. To make this backtracking easier I have added page numbers to the pages which contain the needed information.

The key research questions and arguments are grouped around the conventionalization problem I call the *meaning uncertainty problem* which is introduced in chapter 4. This problem requires agents to bootstrap from scratch a lexical system where meanings show internal structure and are not established a priori or by an external concept formation process.

- I argue that the essence of the two views of lexical learning can and should be captured by two different types of uncertainty or two different conventionalization problems. The first is the problem of mapping uncertainty the other that of meaning uncertainty.

- In Chapter 3 I confirm that competitive cross-situational strategies can successfully solve the mapping problem. This has been shown many times before and the competitive cross-situational strategies I implement are all based on previous work.

- In Chapter 4 I investigate whether competitive cross-situational strategies can tackle the problem of meaning uncertainty and if so in what way? The claim I put forward is that competitive learning algorithms are not up to this task, except when supplied with additional biases.

- I claim that adaptive strategies which (1) do not enumerate competitors, (2) implement flexible processing and (3) show adaptive alignment taking into account feedback from flexible processing can solve the meaning uncertainty problem in a more convincing and natural manner (Chapter 5).

- I argue that adaptive strategies are compatible with important tenets from cognitive grammar and models certain psychological phenomena such as entrenchment, semantic erosion, extension, comparison and categorization. Adaptive strategies also subscribe to a usage-based perspective of language learning and language change. Later I even argue that it shows remarkable parallels with the theory of grammaticalization.

- I investigate the properties of the emergent lexical systems coming from adaptive strategies and show that they capture certain phenomena also found in human natural languages. (Chapter 5 and 6)

- I investigate whether semantic categorization in grammatical constructions can also emerge through adaptive strategies. My claim is that they can and I show support for this claim in Chapter 7.

- I argue for a different approach to classifying language games. Instead of taking the interaction script as the definitional feature, I defend the position that conventionalization problems are a better candidate for classification. All language games in this thesis embody this approach but probably chapters 2 and 3 make this point most clearly.

As part of this thesis I took the effort of reimplementing and sometimes re-interpreting past strategies, especially in Chapters 2, 3 and also 4. I found it important and illuminating to not solely focus on my own contributions but present them as extensions or contrast them with previously proposed strategies. As such I hope that the thesis can serve as an overview and consolidate a larger number of research findings within the agent-based modeling approaches to language evolution[1]. Reimplementing these past strategies also facilitates comparisons since different strategies, both old and new, can be tested on the exact same data sets. This allowed me to validate past results and claims and could sometimes uncover previously unknown limitations or strengths by applying the strategies on new kinds of data.

Related to this, the thesis contributes to a new methodology in the study of evolutionary linguistics where concepts from evolutionary biology, cognitive linguistics and computer science are brought together. Computer science provides tools and design patterns, cognitive linguistics contributes the necessary insights that serve as inspiration for the reported language strategies, and evolutionary biology hints towards an explanatory framework. Only when the disciplines contribute to and from each other has the methodology truly succeeded.

[1]The overview however is far from exhaustive and is still selected in such a way as best support the arguments I wish to convey.

1.6.1 Chapter 2: Strategies for Naming Games

Chapter 2 presents and compares different approaches to the Naming Game, the most well-understood of all language games. It is still part of the first introductory part of the thesis because it was mainly motivated by methodological concerns. It employs the most important concepts used throughout this thesis in the context of a very easy to understand language game.

The first issue addressed in this chapter is trying to find a definition for the Naming Game. After almost two decades of research in the language game paradigm many games have been called Naming Games. I show in this chapter that there is no common definitional "essence" to be found among these games, which forced me to question what could or should be the best candidate for defining language games. As I have explained earlier I will not follow the more traditional approach of defining a game through the interaction script, but instead the definitional aspect is the conventionalization problem it evokes, regardless of the script.

In this line of thinking I link the Naming Game to one particular conventionalization problem, that of name or word form competition. The essence of the problem is that different (competing) words or names are introduced to refer to the same referent. The goal however is to arrive at a single shared word per referent. For the remainder of this thesis it is crucial to thoroughly introduce this problem since any multi agent setup where agents engage in local pairwise interactions will have to deal with this problem. Even the adaptive strategies introduced in the fifth chapter still require the addition of a word form competition strategy.

The chapter also serves the overarching objective of consolidating previous work. Most strategies presented in this chapter are not new, but they have never been compared to each other before in an identical experimental setup. In order to structure these strategies I distinguish between minimal and non-minimal strategies. Minimal strategies are those that do not implement a scoring mechanism to solve the conventionalization problem, non-minimal strategies do.

Baseline Naming Game Strategy: This strategy only serves the purpose of illuminating the problem of form competition as it does not implement an alignment strategy. It allows agents to reach communicative success but does not solve the conventionalization problem introduced by the Naming Game. This strategy has not been presented or discussed in previous work since it essentially fails at successfully playing the Naming Game.

Imitation Strategy: When implementing and writing about this strategy I thought it was a new original strategy not yet discussed before in the

literature on the Naming Game. I found out, however, that around the same time Baronchelli *et al.* (2011) had published on this particular strategy. I believe it is the most elementary strategy that can solve the problem of form competition. As opposed to the Baseline NG Strategy it implements a winner-takes-all dynamics which allows it to solve the problem of word form competition.

Minimal NG Strategy: This strategy is now a very well-known strategy for the Naming Game and it was first introduced by Baronchelli *et al.* (2006a).

Lateral Inhibition strategies: Lateral inhibition strategies extend the dynamics of the Minimal NG Strategy with a more subtle scoring scheme. Agents can keep track of competing names for a longer period of time. Chronologically lateral inhibition strategies were introduced before any of the above minimal strategies. Examples of these strategy can be found in (Vogt & Coumans, 2003; De Beule & De Vylder, 2005; Steels & Belpaeme, 2005).

Frequency Strategy: The frequency strategy was most likely the very first Naming Game strategy. In the context of Naming Games it was already used by Steels & McIntyre (1999).

One of the main open research questions, that also motivated me to reimplement these strategies, is to what extent the addition of a scoring mechanism improves upon the Minimal NG Strategy. These strategies have been investigated in isolation but not yet in a direct comparison. The hypothesis was that non-minimal strategies (i.e. lateral inhibition and frequency strategies) would show a dramatic improvement with respect to speed of convergence.

Who should read this chapter and how?

If you are new to the language game paradigm it is suggested to read the chapter in full. It will clarify the types of problems that are dealt with, how a language game can evoke such problems, and how different strategies can solve the same problem.

Even for people with a background in language games it might still be worthwhile to read certain parts. For example, the way in which I define games might be new and bringing together all these strategies has not been done before.

With regard to the remainder of the thesis it is only important to understand the problem of form competition and the Interpolated Lateral Inhibition Strategy (Section 2.4) as this strategy will return quite often.

1.6.2 Chapter 3: Cross-situational strategies in Minimal Guessing Games

Chapters three to six revolve around the problem of uncertainty in word meaning. In this thesis I distinguish between two approaches to lexical learning and bootstrapping, which leads to two different conventionalization problems.

The first problem, which I call the problem of mapping uncertainty, is investigated in Chapter 3. The problem of mapping uncertainty is finding one-to-one mappings from a set of forms to a set of (atomic) meanings when at first exposure the mapping that should be made is unknown. The mapping problem has been heavily investigated in a computational setting, especially through a learning strategy known as cross-situational learning Siskind (1996). In the same line as the Naming Game chapter I introduce a language game, which I call the Minimal Guessing Game, which aims to evoke only the additional problem of mapping uncertainty.

In this chapter I present one non-cross-situational strategy and three cross-situational strategies. Again the goal is not to present novel strategies, all three cross-situational strategies are based on previous work (see overview of strategies below). The aim of this chapter is to introduce the problem of mapping uncertainty, to define a language game that evokes this problem in a minimal way, and to present the reader with an overview of which strategies have been proposed to tackle this problem. In short, cross-situational strategies explicitly *enumerate* competing meanings for the same word and rely on a selectionist competition dynamics to "find" the conventional meaning (i.e. the correct mapping), which is why I call such strategies *competitive* cross-situational strategies. The strategies discussed in this chapter are:

Baseline GG Strategy: This non-cross-situational strategy is primarily supplied to show that cross-situational statistics are in fact crucially important. A strategy that does not track such statistics cannot scale well along multiple dimensions.

Set-Based CS Strategy: This strategy implements cross-situational learning in a minimal way, without keeping statistical scores. It was introduced by Smith *et al.* (2006) and further investigated by De Vylder (2007) and Blythe *et al.* (2010).

Frequency CS Strategy: In this strategy agents can keep track of the co-occurrence frequencies between words and meanings. This allows them to better track the cross-situational statistics and makes them more robust to inconsistent input. A Bayesian variant of the Frequency CS Strategy was previously introduced by Smith (2003a) and Vogt & Coumans (2003).

Flexible CS Strategy: Another cross-situational strategy introduced by De Beule *et al.* (2006). It tries to make the agents more responsive to new potentially emerging conventions.

Who should read this chapter and how?

If you are new to (traditional) cross-situational learning and want to know more about this type of learning and the mapping problem then I advise to read the full chapter. If cross-situational learning is new to you but you are mainly interested in the adaptive strategies coming later it is most likely enough to only read the introductory part.

If you are acquainted with traditional cross-situational learning then this chapter can be skipped. However, I would suggest to nevertheless read the introductory part of the chapter since it outlines the conventionalization problem of mapping uncertainty and introduces the Minimal Guessing Game.

1.6.3 Chapter 4: Competitive strategies for Compositional Guessing Games

This chapter introduces the Compositional Guessing Game and the meaning uncertainty problem. In the Compositional Guessing Game meanings show internal structure and utterances can consist of multiple words. I have chosen for the simplest type of meaning structure, namely a set. In the Compositional Guessing Game both meanings and objects are represented as sets of attributes.

This representational change leads to a different conventionalization problem since agents need to establish and align the meanings (sets of attributes) of their words as well. The main problem is no longer one of finding a correct mapping between a set of forms and a set of meanings, but to construct and align the meanings themselves through communication. I call this problem the *meaning uncertainty problem*.

The goal of this chapter is to investigate whether competitive cross-situational strategies as proposed in Chapter 3 can scale to the meaning uncertainty problem. Although I introduce two strategies in this chapter,

their key learning algorithms are the same as those of the Frequency CS Strategy of the previous chapter.

Baseline Competitive Strategy: This strategy is simply the Frequency CS Strategy from the Chapter 3 ported to cope with structured meanings. It holds true to the core aspects of a competitive strategy, which is an enumeration of hypotheses coupled with a subsequent pruning of this list.

Discriminative Strategy: This strategy extends the Baseline Competitive Strategy by adding an extra pre-processing component. In this case the component allows agents to find discriminative subsets of the topics they need to communicate about.

Based on the experimental results, I argue that competitive strategies either show severe scaling problems (the Baseline Competitive Strategy) or tend to be biased toward certain, most likely atomic, lexical systems (the Discriminative Strategy).

I thus conclude that competitive strategies of cross-situational word learning are not easily extended to tackle the problem of meaning uncertainty. I claim that these limitations stem from the way in which these approaches represent and cope with uncertainty.

Who should read this chapter and how?

The first few sections (until and including Section 4.2) are crucial in order to understand the remainder of the thesis as it introduces the Compositional Guessing Game and the problem of meaning uncertainty. This game and problem plays a pivotal in the remainder of the thesis. Section 4.5.2 is also important if you are interested in the details of how structured objects are generated. The object generation algorithm explained there is also used in the subsequent chapter.

In Section 4.5.1 I argue why we should not only be interested in communicative success and lexical alignment but also should investigate the characteristics of the emergent lexical systems. I argue that we should look for strategies that lead to usage-based lexical systems. This is a fundamental point which I will return to in the following chapters as well.

The two strategies that are introduced are not new. I am not the first to introduce them and they stay true to the core principles of the strategies introduced in Chapter 3. It is not required to understand these strategies in order to understand the adaptive strategies introduced later. It does help to understand *why* adaptive strategies are proposed as they are and to see the crucial difference between both approaches.

1.6.4 Chapter 5: Adaptive strategies for Compositional Guessing Games

In chapter 5 I first try to clarify why competitive approaches to learning show the problems discovered in Chapter 4. The principal reason lies in the manner in which competitive strategies represent and deal with uncertainty. Competitive strategies explicitly represent competing hypothesis and rely on a selectionist elimination-based learning to find the target representation. When confronted with large meaning spaces such a search algorithm is too weak.

Competitive approaches are, however, more attuned to a view of word learning as mapping. This view essentially takes the meanings (or concepts) to be readily available and all that remains is mapping words to those apriori established meanings. In this case, as shown in Chapter 3, competitive approaches make sense and solve the problem (which is the mapping problem) rather convincingly.

In the Compositional Guessing Game the problem of word learning is not just viewed as mapping but as a problem where the meanings themselves need to be established. I introduce an approach, which I still believe is cross-situational, but that tries to steer clear of the assumptions made by competitive approaches. This approach, which I call adaptive as opposed to competitive, has three definitional characteristics:

1. No enumeration of competitors. No selectionist competition dynamics. At all times only a single hypothesis is maintained per word.

2. Flexible processing (both in production and interpretation) which stresses re-use of existing meanings even when they are not fully compatible with the intended meaning.

3. Internal adaptation of meanings based on the feedback from flexible processing.

In this chapter I introduce two adaptive strategies and show whether and how they can tackle the problem of meaning uncertainty. I investigate the characteristics of the emergent lexical systems and the impact of different important parameters such as population size, attribute size, etc. Also the impact of certain cognitive operations such as mutual exclusion is investigated.

Who should read this chapter and how?

This the pivotal chapter of the thesis, introducing the adaptive approach and showing how it performs in the Compositional Guessing Game for the problem of meaning uncertainty.

If there is only one chapter you read from this thesis, this is the chapter to read. Although bear in mind that you might need to backtrack to some parts of Chapter 4 (see previous section).

This chapter is also crucial to understand any of the following chapters. Both chapters 6 and 7 rely on an understanding of the Weighted Adaptive Strategy. But the Weighted Adaptive Strategy largely relies on the Baseline Adaptive Strategy thus they should both be understood, until progressing further in the thesis.

As I argue in the conclusion of the chapter I believe adaptive strategies are highly compatible with core tenets of usage-based cognitive linguistics. Therefore I hope that readers with a background in cognitive linguistics or psycholinguistics would, if they read only one chapter, read this chapter. I believe adaptive strategies supply them with a computational model that also shows the robustness of their view of language.

I would also love to see experiments in the field of experimental semiotics trying to find credence for adaptive learning. For example trying to focus on the nature of the initial hypothesis formed by the subjects and then analysing how this hypothesis is changed throughout experimental trials.

1.6.5 Chapter 6: Robustness in Compositional Guessing Games

In Chapter 6 I further flesh out adaptive strategies using robotic experiments. The use of robots introduces a specific problem and stresses robustness of the language strategies.

More specifically it introduces a more difficult variant of the meaning uncertainty problem investigated in Chapters 4 and 5. The difficulty arises from not having exactly the same experiences of the objects communicated about. In the experiments this arises from the robots taking different positions in the room but it could also arise out of noisy conditions or simply having different conceptualizations of the context.

I first investigate whether adaptive strategies can deal with this more difficult variant and if so how they achieve this. This also leads to a deeper investigation of the alignment dynamics that the adaptive strategies exhibit, a dynamics I call *shaping dynamics*.

Finally I also compare a competitive strategy with an adaptive strategy in this chapter.

Who should read this chapter and how?

This chapter further fleshes out the dynamics and properties of adaptive strategies so also counts as an essential part of the thesis. It does not introduce a new strategy but for first time compares competitive and adaptive approaches.

The additional challenge that is solved in this chapter is also part of the problem addressed in the following chapter. In order to understand how that challende is solved it is thus necessary to have read this chapter.

1.6.6 Chapter 7: Adaptive strategies in the domain of grammar

In all previous chapters we have been focusing on lexical problems and the emergence of word meanings. In my view, adaptive learning is not constrained to only the lexical level at all. In this chapter I investigate whether the underpinnings of adaptive strategies can be extended beyond the scope of meaning uncertainty and lexical systems. As a case study I show that a word-order pattern based on semantic categorizations, inspired by adjectival orderings found in a large number natural languages, can be bootstrapped and aligned by a population of agents employing the same underlying ideas employed by the previously introduced adaptive strategies.

I dissect the problem of bootstrapping and aligning a category-based word order pattern to exhibit two additional problems on top of the problem of meaning (in this case category) uncertainty. The first is that alignment is no longer based on experiences but on word meanings and these word meaning might not be fully aligned within the population. The second, and more crucial problem, is that in such a pattern (and in much of grammar), categories can be skipped if the speaker does not wish to express those features. This problem, which stems from the optionality of slots, introduces a difficult and new challenge to the alignment dynamics of adaptive strategies.

Who should read this chapter and how?

This chapter relies very heavily on Chapter 5 and to a lesser extent also on Chapter 6 so these are more or less required reading.

For construction grammarians this chapter is interesting because it shows how the same type of strategies can be extended to shape not only lexical constructions but also larger constructions. Construction grammarians believe that both lexical and grammatical constructions are part of one continuum and thus sharing the same alignment strategy supports the credibility of this view.

For computational modelers in the evolution of language and grammar this chapter presents a particular view of how semantic categories might emerge out of lexical material. And it shows that agents do not need to represent competing word orders to bootstrap an aligned word order pattern.

1.7 Overview of publications

Not all the work that I have conducted could be bundled in this thesis. I have decided for this thesis to focus mainly on lexical systems and leave out most of the work I have done on grammar. In this section I would like to touch upon the peer-reviewed publications that have been published during my time as a PhD student.

1.7.1 Earlier work on the emergence and formation of grammar

The first two publications which I co-authored with my supervisor Luc Steels were published in my first academic year at the VUB AI-Lab and both focused on the emergence of grammatical systems. The investigated hypothesis was to what extent linguistic search during processing could count as a trigger for the emergence of grammatical constructions, a hypothesis proposed a year earlier by Steels (2005).

For Steels and Wellens (2006) I implemented a set of diagnostics and repair strategies in Fluid Construction Grammar (FCG) (Steels, 2011a) that operationalize this hypothesis and show that grammar can indeed be used to minimize search during parsing. The diagnostics and repair strategies ranged from lexical to grammatical, allowing the agents to first bootstrap a lexicon and later, upon activating grammatical learning strategies, a grammatical system.

One drawback of the models in this paper was that the diagnostics and repairs had to be activated by the modeler himself. The agents could not self-monitor their competence in order to move to new levels of linguistic complexity. In Steels and Wellens (2007) I further refined the model to incorporate the autotelic principle (Csikszentmihalyi, 1990; Steels, 2004a). This principle allows the agents to monitor their competence and communicative success. Using this monitoring they could gradually move between different levels of linguistic complexity.

Although primarily the work of Remi van Trijp, I also co-authored and co-developed a model for multi-level selection in the emergence of language systematicity (Steels, van Trijp and Wellens; 2007). In this work we forward

and supports through a computational model the claim that in order to arrive at a systematic language system, agents require an interconnected scoring dynamics across linguistic constructions. Constructions need to be represented in a network structure, connecting constructions according to e.g. part-whole relations. The updating dynamics during alignment need to follow these edges in the network and also adjust the scores of neighbouring constructions.

- 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* (p. 7688). Berlin.

- Steels, L., & Wellens, P. (2007). Scaffolding Language Emergence Using the Autotelic Principle. In *IEEE Symposium on Artificial Life 2007* (Alife 07) (p. 325332). Honolulu, HI.

- 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) (p. 421434). Berlin.

1.7.2 A move toward adaptive models for the emergence of lexical systems

One limitation I was unsatisfied with in my earlier work on grammatical strategies and systems was that the learning strategies could not scale to large hypothesis or meaning spaces. The strategies would break down when the uncertainty became too large. In the domain of lexical learning there seemed to be ongoing work regarding these scaling issues so I turned to lexical strategies and systems for inspiration in order to investigate more robust models of learning. This lead to most of the work presented in this thesis.

Since the work addressed in those publications is covered deeply in this thesis I only present a listing of the related publications.

- Wellens, P., Loetzsch, M., & Steels, L. (2008). Flexible Word Meaning in Embodied Agents. *Connection Science*, 20, 173191.

- Wellens, P. (2008). Coping with Combinatorial Uncertainty in Word Learning: A Flexible Usage-Based Model. In A. D. M. Smith, K. Smith, & R. Ferrer-i-Cancho (Eds.), *The Evolution of Language. Proceedings of the 7th International Conference* (EVOLANG 7) (p. 370377). Singapore.

- Wellens, P., & Loetzsch, M. (2012). Multi-Dimensional Meanings in Lexicon Formation. In L. Steels (Ed.), *Experiments in Cultural Language Evolution* (p. 143-166). Amsterdam.

1.7.3 Grammatical processing: dependency networks

In a later stage I returned to the domain of grammar. The first goal was to incorporate and test the insights from adaptive strategies into the grammatical investigations. This undertaking is still a work in progress and largely future work.

At that time we had different people using Fluid Construction Grammar with large scale grammars and a pressing problem was to speed up the processing. In (Wellens, P. and De Beule J. 2010; Wellens, P. 2011) I investigated cognitively plausible models in which the FCG processor could infer which grammatical constructions showed dependencies among each other. Encoding these dependencies leads to a usage-based dependency network structure of the complete inventory of constructions, both lexical and grammatical. Such a dependency network could then be used during processing (both parsing and production) to dramatically improve speed of processing by priming (prioritizing) the most likely constructions.

I have also co-authored two recent publications regarding linguistic processing in Fluid Construction Grammar. In (Steels, De Beule and Wellens, 2012) we show how FCG can be hooked up to real robots and used in grounded experiments. Different parts of the constructional FCG representations become hooked to the embodied data through another formalism for embodied semantics. In (Beuls, van Trijp and Wellens, 2012) we propose a dual- or meta-layer architecture for linguistic processing which distinguishes between routine processing and "special" meta-layer processing. In this meta-layer, diagnostics continuously monitor the routine layer to check for potential problems. These problems are then reported to repair strategies that can try to alleviate these problems by introducing new linguistic items or shifting the usage pattern of others.

- Wellens, P., & De Beule, J. (2010). Priming through Constructional Dependencies: a case study in Fluid Construction Grammar. In A. D.M. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.), *The Evolution of Language* (EVOLANG8) (p. 344-351).

- Beuls, K., & Wellens, P. (2010). Linking Constructions and Categories: A Case Study for Hungarian Object Agreement. In *Book of Abstracts of the Sixth International Conference on Construction Grammar* (p. 4445).

- Wellens, P. (2011). Organizing Constructions in Networks. In L. Steels (Ed.), *Design Patterns in Fluid Construction Grammar* (p. 181-202). Amsterdam.
- Steels, L., De Beule, J., & Wellens, P. (2012). Fluid Construction Grammar on Real Robots. In L. Steels & M. Hild (Eds.), *Language Grounding in Robots*. Berlin.
- Beuls, K., van Trijp, R., & Wellens, P. (2012). Diagnostics and Repairs in Fluid Construction Grammar. In L. Steels & M. Hild (Eds.), *Language Grounding in Robots*. Berlin.

Chapter 2

Strategies for Naming Games

The concept of language games was introduced by the 20th century philosopher Ludwig Wittgenstein in his philosophical investigations (Wittgenstein, 1967). In this work he provides what must be the most famous of all language games, which is, perhaps not by accident, about building.

> Let us imagine a language ... The language is meant to serve for communication between a builder A and an assistant B. A is building with building-stones; there are blocks, pillars, slabs and beams. B has to pass the stones, and that in the order in which A needs them. For this purpose they use a language consisting of the words 'block', 'pillar', 'slab', 'beam'. A calls them out; –B brings the stone which he has learnt to bring at such-and-such a call. – Conceive this as a complete primitive language. (Wittgenstein, 1967, passage 2)

Wittgenstein did not believe this is the way language works. In fact he provided this example to criticise a certain conception of language (that of Augustine). He saw language games much more as a tool, used to shed light on some aspect of a linguistic interaction. Sometimes these aspects can be brought out through similarities to real world language use but often (and especially in the earlier language games) through their dissimilarities. The Naming Game, which is the language game discussed in this chapter, also bears little similarity to real world language use. Language games should not be taken as representative of even a part of language, or as Wittgenstein put it, as if they were first approximations, ignoring friction and air-resistance (Wittgenstein, 1967, passage 130).

From the mid nineties computational implementations of increasing complexity of language games have been proposed. They are known as *Naming Games* (Steels & McIntyre, 1999), *Guessing Games* (De Beule *et al.*, 2006),

Description Games (van Trijp, 2008b) and so forth. The game discussed in this chapter, the Naming Game, very much resembles the game described by Wittgenstein in the quote above.

2.1 Definition and history of the Naming Game

Although the Naming Game is arguably the most restrictive language game, it was not the first language game to be operationalised. Trying to determine the very first reported Naming Game implementation turned out to be quite a challenge. For example the very definition of a Naming Game has undergone quite some change over the past fifteen years.

In this paper the focus is on the *Non Grounded Naming Game*, all experiments are done in simulation on computers without the agents being embodied in robots. The Grounded Naming Game (Steels & Loetzsch, 2012) shares only limited features with its Non Grounded variant. In all the following whenever a reference to *the* Naming Game is made, it is meant to be the Non Grounded Naming Game. More often than not statements would be false if the Grounded version was to be included.

Instead of presenting a strict definition of the Naming Game I will start with presenting features found in many but not necessarily all Naming Games.

1. A Naming Game involves two participants, a listener and a speaker.
2. In a Naming Game a speaker utters a single name to refer to one object in a context.
3. Objects do not show internal complexity and names can only refer to objects as a whole and not to categories or features.
4. At the end of the game both participants know the intended referent and can thus unambiguously make a pairing between name and object. Uncertainty about the referent of a name does not or only marginally occurs.
5. Different names for the same referent (form competition) can be introduced in the population and increase with its size.

With the aid of this list it is possible to map out a short history of the Naming Game and define the game as it will be used in this chapter. The very first operationalized language games (Steels, 1995, 1996), although referred to as Naming Games in (Steels & Kaplan, 1998a), do not show many of the features above. In Steels (1996) the game is played by more than two participants, there are multiple referents (here called meanings instead of

objects) per game and the referent of a single word is ambiguous. In Steels (1995) there is a single referent (which is an object in a context), but this object is described through spatial terms and thus shows internal complexity. Words do not refer to objects but to spatial terms. These games miss crucial constraints essential to still be categorised as a Naming Game today.

Only two years later in (Steels & Kaplan, 1998a,b; Steels, 1998a; Steels & McIntyre, 1999) a game is introduced, explicitly called a Naming Game, which shares features 1,2,3 and 5 of the above list. The setup in these games does allow uncertainty at the end of the game through different variants of "noise" parameters or interpretation probabilities, as such invalidating the fourth constraint. In 2005 and 2006 the Naming Game got renewed attention from a far more theoretical perspective (Lenaerts *et al.*, 2005; De Vylder & Tuyls, 2006; Baronchelli *et al.*, 2006b). It was only in these papers that feature 4 (no uncertainty) became standard part of the Naming Game. De Vylder & Tuyls (2006) and Baronchelli *et al.* (2006b) argued that when features 3 and 4 are present the script of the game can be simplified. There is no need to simulate multiple objects per game, one object suffices. They went even further and showed that, without loss of generality, there doesn't even have to be more than one object at all. Whether the agents have to agree on names for 1 or $n \geq 1$ objects has no impact on the dynamics when feature 3 and 4 are present. This more theoretical approach to the Naming Game has been investigated in depth over the past few years (Lu *et al.*, 2008; Liu *et al.*, 2009; Lipowski & Lipowska, 2008; Baronchelli *et al.*, 2008).

The Naming game as discussed in this chapter follows the approach of De Vylder & Tuyls (2006) and Baronchelli *et al.* (2006b). All five features (or restrictions) are present. In this thesis the script of the game is kept closer to that of Steels & Kaplan (1998a) with multiple objects per context so that it is more similar to that of games introduced in later chapters. To keep simulations fast and since it doesn't impact the results in a significant way the number of objects is kept small at five.

Also note that in a Grounded Naming game only features 1 and 2 tend to be present, bringing with it all kinds of problems not present in the Naming Game as discussed in this chapter.

2.2 Script of a Naming Game

The central question in a Naming Game experiment is how a population of agents can bootstrap and maintain a shared set of names for a set of objects. In a Naming Game experiment the agents can only engage in Naming Games, which are local to the participating agents only. The agents cannot negotiate names in any other way. Let us start by taking a closer look at the script of

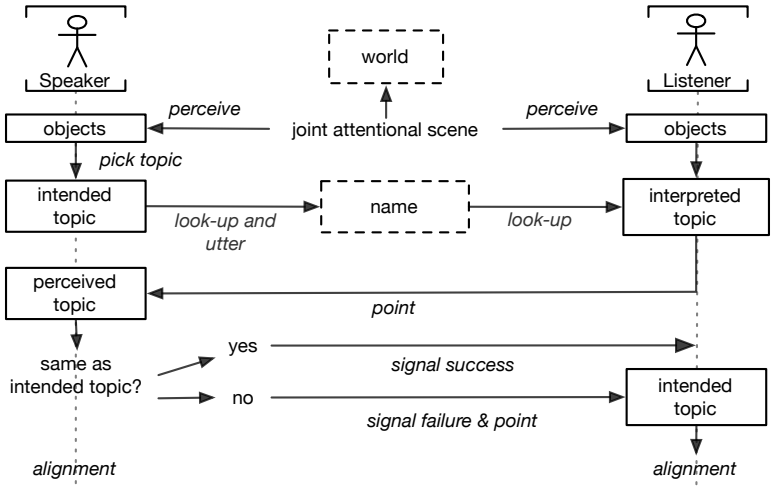


Figure 2.1: Script of a Naming Game. Squared boxes represent data entities. Italicised text refers to processes. The look-up processes can fail and therefore have diagnostics and repair strategies associated with them. A Naming Game experiment consists of thousands of these interactions.

a single Naming Game.

The Naming Game as discussed in this chapter requires objects $o_i \in O$ and agents $a_1 \cdots a_n \in P$ (population). For every game a context is generated and two agents a_s and a_l are chosen from the population. One agent is assigned the role of speaker, the other of listener[1]. Both agents are confronted with the same context C of n objects. When these conditions are met the game can start (also see Figure 2.1).

1. The speaker mentally picks one object from the context (called the *topic*) and utters a name for it.

2. The listener interprets the name and points to the object he believes the speaker intended.

3. The speaker either agrees or disagrees with the listener.

4. Both agents have the opportunity to adapt their inventory of names.

[1]Agents are chosen at random, there is no network topology (see (Baronchelli *et al.*, 2007; Van Segbroeck *et al.*, 2010) for more on the role of network topologies).

The script can be made less verbose without any loss of generality. For example it is not required that the listener points to the object since per definition (feature 4) it is impossible that the listener would have associated a different meaning to the name.

On the *representational* level not many options are available in strategies for the Naming Game. The nature of the game requires meanings to be represented as non-compositional concepts. Likewise the form of the names is also assumed to be represented as non-modifiable strings[2]. Each agent a needs a bi-directional memory M_a pairing names n_j with objects o_i. One option left open by the specification of the Naming Game is whether agents can score each name-object pairing. The idea is that maintaining such a score helps the agent to better determine the communicative outcome when using that particular name. The strategies discussed in this chapter will differ in this respect.

What kind of language strategy do the agents need to share in order to arrive at the system of a shared set of names? A language strategy needs to supply both representational and processing aspects.

2.3 Minimal strategies for the Naming Game

In a Naming Game experiment a population of agents should develop a *shared* and *minimal* set of names for the set of objects. Minimal means there is only one name used for each object and all agents share this name.

In this section we investigate three *minimal strategies*, so called because they keep representational and processing capabilities as minimal as possible. Most prominently, none of the minimal strategies implements a scoring mechanism.

2.3.1 The Baseline NG Strategy

Since the Baseline NG Strategy is a minimal strategy it does not allow agents to score their name-object pairings. All that is assumed here is that an agent can store pairs $< n, o >$ in his memory M_a. Such a memory can be implemented in many different ways, ranging from a simple set of pairs to a hash-table or more structured inventories.

For baseline processing each agent a uses two look-up functions $f_{produce}(o, a)$ and $f_{interpret}(n, a)$ which access the memory M_a of agent a and look-up an

[2]In all the experiments in this thesis word forms are assumed to already be segmented and holistic in nature. These strings are transferred to the other agent without relying on speech. A long tradition of computational simulations does exist for the domain of speech (de Boer, 2000; Oudeyer, 2005) [see (de Boer & Fitch, 2010) for an overview].

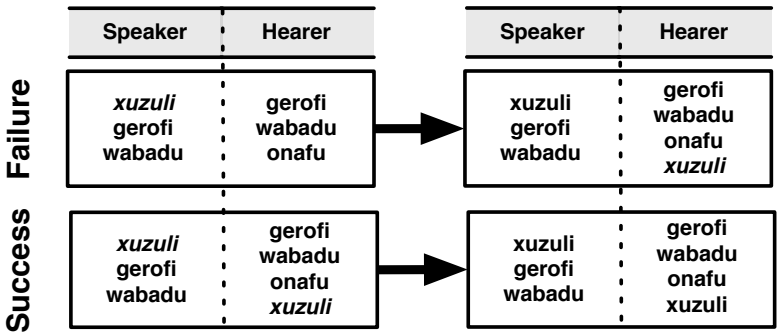


Figure 2.2: Update mechanics for the Baseline NG Strategy. Only the names for the topic object are shown. From a list of available names the speaker utters "xuzuli". In case of failure the listener does not know the name and adopts the new name. In case of success agents do not update their inventories.

object o or a name n respectively. Since in our flavor of the Naming Game uncertainty about meaning is absent $f_{interpret}(n, a)$ returns either one object in case the agent already encountered the word once or no object if it is the first exposure. The production function $f_{produce}(o, a)$ however returns a set of names, which can be either empty, have one element, or multiple elements depending on whether several names are competing for the same object. In case multiple names are returned the agent will, before uttering, pick a random name from this set of competitors. The implementation of these two functions depends on the implementation of the memory architecture M_a and should not impact the dynamics of the strategy as long as it conforms to the input output behaviour just explained. The goal of the Naming Game is that agents agree on a *single* preferred name per object. This means all agents should always return the same name from $f_{produce}(o, a)$.

All agents start out with empty inventories and thus require a capacity to introduce a new name if needed. When the speaker does not yet know a name for his chosen topic t he *invents* a random name n[3], links it to topic t and stores the pairing $< n, t >$ in his memory M_a. As listener, an agent needs a similar capability to store or *adopt* new names. When $f_{interpret}(n, a)$ fails to return an object o the listener first waits for the speaker to point to the intended topic t and adds the association $< n, t >$ linking name and

[3]The algorithm guarantees that the same name cannot be invented twice ruling out "accidental" homonymy or meaning competition.

object. Production and interpretation are thus augmented with two learning functions $f_{invent}(t, a)$ and $f_{adopt}(n, t, a)$.

The two processing functions $f_{produce}$ and $f_{interpret}$ and the two learning functions f_{invent} and f_{adopt} make up the Baseline NG language strategy for the Naming Game. Figure 2.2 schematically shows the updating mechanics just explained.

2.3.2 Measures for naming strategies

In evaluating strategies for the Naming Game we use the following measures:

Communicative success: A Naming Game is communicatively successful when the listener points to the correct object. Due to the restrictions of this game, when a listener knows a name he will point correctly, thus communicative success can be simplified to whether the listener knew the spoken name (i.e. $f_{interpret}(n, a)$ returns an object). In the graphs a running average of 25 is used.

Name competition and alignment: With name competition the average number of competing names per object is measured. Name alignment measures the level of alignment towards one preferred name for each object. Given the set of all expressible objects O, the population P and the production function $f_{produce}(o, a)$ which returns the set of names that agent $a \in P$ prefers for object $o \in O$.

$$\text{Name competition} = \langle | \bigcup_{a_j \in P} f_{produce}(o_i, a_j) | \rangle_{o_i \in O} \tag{2.1}$$

$$\text{Name alignment} = \frac{1}{\text{Name competition}} \tag{2.2}$$

In a fully aligned population, where all agents prefer the exact same name for each object, name alignment equals 1.

Alignment success: This measure combines communicative success and name alignment. It is, like communicative success, a Boolean value which is only true when the listener knows the name spoken by the speaker (i.e. communicative success) and it is also his own preferred name for that object (if he were speaker he would utter it as well). In other words $f_{produce}(t, a_{speaker}) = f_{produce}(t, a_{listener})$ and $|f_{produce}(t, a)| = 1$ for both agents.

All Boolean measures, like communicative and alignment success, are plotted as running averages over a window of 25 games. In what follows I

will sometimes talk of *full* Communicative success or full Alignment success (or any other measure) to denote that the population has reached 100% for that measure.

All the above measures only depend on the output behaviour of $f_{interpret}$ and $f_{produce}$ and not on the internal representations of the agents, which makes them independent of a particular strategy.

2.3.3 Experimental results of the Baseline NG Strategy

We are now in a position to run the Baseline NG Strategy and see how it performs. Unless otherwise stated all following Naming Game experiments are run with a population of 50 agents and a total of 5 objects. As explained earlier this number does not impact in a significant way the dynamics of the strategies. The only difference is that reaching coherence takes slightly longer since instead of a single name the agents have to establish five names. Context size is irrelevant in the type of Naming Game that is discussed in this chapter. In most graphs (e.g. Figure 2.3) the x axis represents the number of games played either per agent or in total. With a straightforward calculation one can switch between total number of games or games per agent.

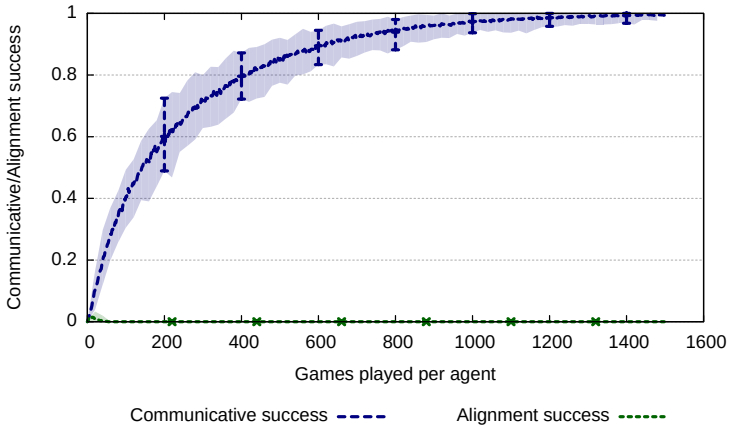
$$\text{games played per agent} = 2 \times \frac{\text{total number of games}}{\text{size of population}} \tag{2.3}$$

Results are averaged (mean) over 100 experimental runs with error bars showing the 5 and 95 percentile[4].

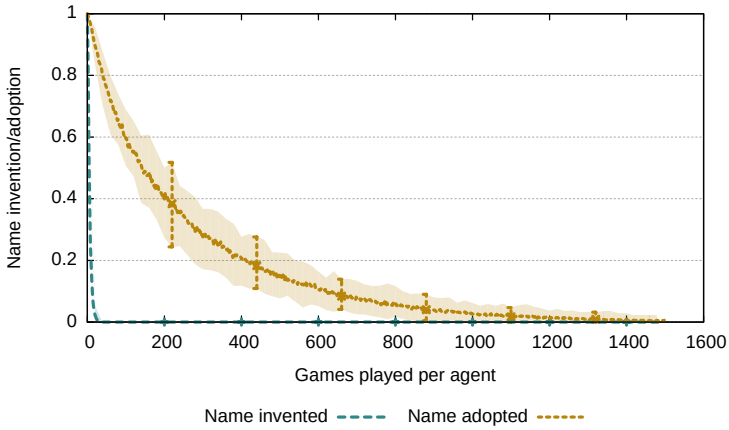
Figure 2.3a shows that the Baseline NG Strategy is sufficient for a population to reach full communicative success. Full communicative success entails that the agents know all names used within the population, which in turn means that invention and adoption must have come to a halt which is indeed the case as shown in Figure 2.3b. In the baseline strategy invention stops as soon as each agent has played one game about each object. Adoption takes much longer because every name that was ever invented has a random chance to be spoken for its associated object and thus in the long term will get adopted by all agents. Also note that for a population of n agents, a single invention requires $n - 1$ adoptions.

Although the population reaches communicative success the emergent language system is wanting with respect to alignment success. The agents invent and need to maintain a large set of names per object. Each game is local

[4]The error is shown either by bars or by a transparent filled curve surrounding the graph, everything inside this filled curve is in the 5 to 95 percentile. An example of both is shown in Figure 2.3.

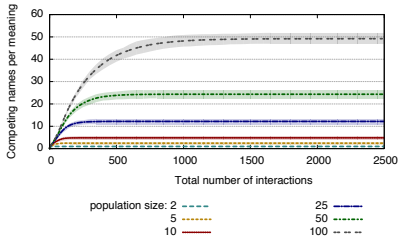


(a) Communicative and alignment success

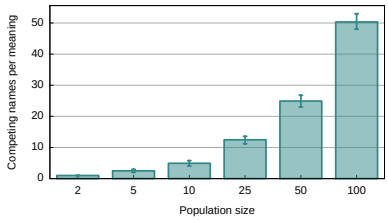


(b) Invention and adoption

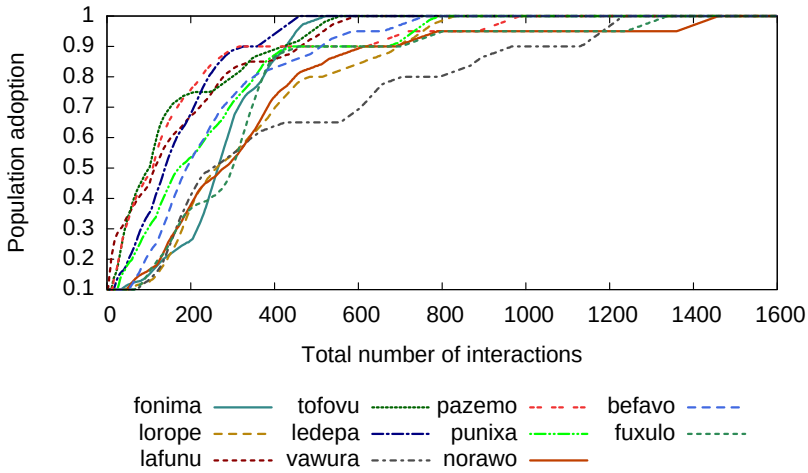
Figure 2.3: Experimental results for the Baseline NG Strategy. (a) Agents reach full communicative success because they store and thus recognize all invented names. Alignment success, however, remains at zero signifying that there is no shared preferred name for each object. (b) Both invention and adoption come to a halt, with adoption taking significantly longer. Parameters of (a) and (b): population size: 50, total object size: 5, number of simulations: 100, error: 5 to 95 percentile.



(a) Evolution of name competitors



(b) Final number of name competitors



(c) Name adoption by population

Figure 2.4: Results for the Baseline NG Strategy without alignment. (a) and (b) The number of competing forms per meaning when the size of the population is increased from 2 to 100. (c) The names floating around for one object o. Each line corresponds to a name n_i and shows the percentage of agents for whom $n_i \in f_{produce}(o, A)$. For sake of clarity, population size was reduced to 20 agents.

to only the two participating agents leading to the introduction of different names for the same object, a problem we call *name* or *form competition*. For example assume there are 10 agents $a_1 \cdots a_{10}$ and in the first game speaker a_3 invents a name n_{first} for object o_5 and transmits it to listener a_7. None of the other agents is aware of the pairing $< n_{first}, o_5 >$. Only if an agent participates in a game with either a_3 or a_7 as speaker and topic o_5 can they be confronted with this name. In the meantime this agent might play games with o_5 as topic, prompting him to also invent or adopt new competing names for it.

As long as agents maintain multiple competing names for an object, without a single shared preference, alignment success is bound to remain low, which is indeed what Figure 2.3a shows. Only in the very beginning minimal alignment is measured but this is only due to small chances that some agents interact and they have only established one and the same name. The baseline strategy thus fails at establishing an aligned naming system in which all agents agree on a single shared name per object.

The problem of name competition increases with the population size because names spread through the population much slower prompting other agents to invent competing names in the process. In fact, the average number of names per object is more or less equal to half of the size of the population as shown in Figure 2.4a and b. Only in the special case when the population consists of two agents do the agents converge on the optimal set of names. Indeed, in this case both agents participate in all games and thus no names can be introduced outside of their knowledge.

Figure 2.4c illustrates this point most clearly. Over the course of 1600 Naming Games we tracked all names in use by the population for a single object. Per name n_i (each line in the graph) the percentage of agents for which $n_i \in f_{produce}(o, a)$ is measured. When a name n reaches 1 for the object o it means that for all agents $n \in f_{produce}(o, a)$. What we see is that for the Baseline NG Strategy all names are accepted by all agents. The strategy does not implement a winner takes all dynamics.

With a bi-directional memory, the processing functions $f_{produce}$ and $f_{interpret}$ and the learning functions f_{invent} and f_{adopt} as outlined for the Baseline NG Strategy a population is capable of reaching a shared lexicon but not a minimal one. We not turn to two minimal strategies that do reach a shared and minimal set of name-object pairings.

2.3.4 The Imitation Strategy

What would be the most minimal adaptation of the Baseline NG strategy that would allow the population of agents to reach communicative success

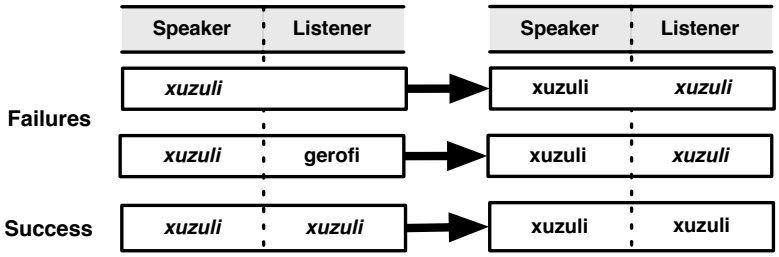


Figure 2.5: Update mechanics for the Imitation Strategy. The speaker utters the word "xuzuli" for the chosen topic. In case of failure the listener adopts the spoken name.

and align on a minimal set of names? In other words how can the population reach full alignment success? Interestingly one solution consists in further *simplifying* the Baseline NG Strategy by disallowing agents to remember different names per object. What if an agent could store only a single name per object?

With this in mind we propose a strategy where a listener, when adopting a new name overwrites any name already stored for the object. The adoption function $f_{adopt}(n, t, a)$ is thus simplified so that instead of extending the list of names for t the agent only remembers the pairing new $< n, t >$ for object t, removing the previous pairing if there was one. The listener will thus imitate the speaker when he has to talk about the same object, which is why we call this strategy the *Imitation Strategy*. Representationally the strategy can be implemented using an even more basic bi-directional memory than the Baseline NG Strategy since only one name per object needs to be remembered. With regard to invention, and look-up it remains the same as the baseline, except that $f_{produce}(o, a)$ can never return more than one name. This strategy corresponds to the Voter Strategy introduced very recently by Baronchelli *et al.* (2011). Its updating scheme is schematically shown in Figure 2.5.

Because of the restrictive nature of the Imitation Strategy the measure for communicative success and alignment success, exceptionally, returns identical results. Indeed when agents reach communicative success it means the listener knows the spoken name, but since agents can only store one name per object it is per definition also his preferred name and thus the agents also reach alignment success.

Experimental results (see Figure 2.6) show that the Imitation Strategy achieves both full communicative success and full alignment success. The

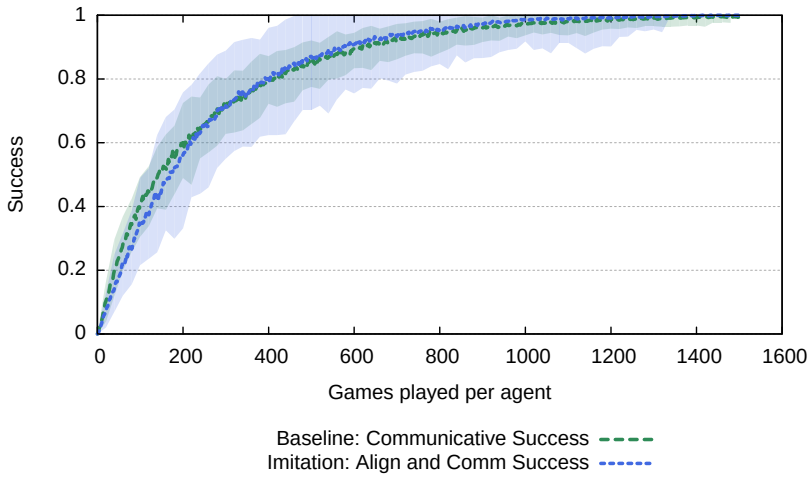


Figure 2.6: Communicative success and alignment success for the Imitation Strategy. The population reaches full communicative and alignment success. Parameters: population size: 50, total object size: 5, number of simulations: 100, error: 5 to 95 percentile.

reason for this success lies in the *winner takes all* dynamics. From the competing names for an object one emerges as the sole preferred name by the whole population as visualised in Figure 2.7. This figure clearly shows the self organizing dynamics of the Imitation Strategy, leading to population-wide preference of the same name per object.

A crucial feature of the Imitation Strategy is the *adaptivity* of the agents in the sense that, as listeners, agents change their own inventory when contradicted by a speaker. Imagine a similar but more "stubborn" variant of the strategy where $f_{adopt}(n, t, a)$ is modified so that in case the listener already knows another name, he does *not* adopt the new pairing $< n, t >$. Agents only adapt their lexicons when they do not yet know a name for the object (see Figure 2.8a). This minor modification to the strategy breaks the winner takes all dynamics and results in populations not capable of even reaching communicative success as shown in Figure 2.8b[5]. A non-adaptive strategy does not yield the desired result because agents are not willing to change their preference.

[5]Only in one exceptional case can a population reach communicative success. For each object the first invented name should spread through the entire population before a second invention for that object takes place.

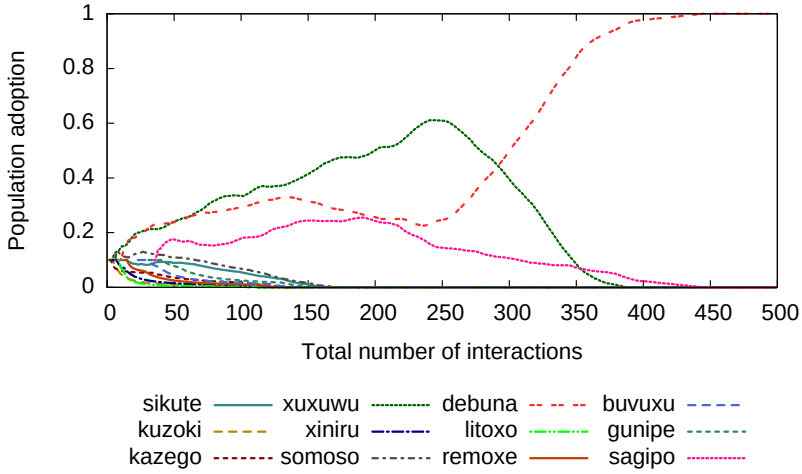


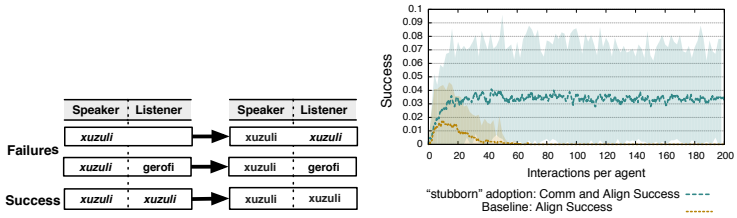
Figure 2.7: Name competition for the Imitation Strategy. Through self-organization the name "debuna" comes out as single preferred name by the full population. A total of 12 names have been invented for the object. Just as for Figure 2.4(c), a population of 20 agents was used. Comparison to this figure clearly illustrates the difference in dynamics.

2.3.5 The Minimal NG Strategy

The last minimal strategy discussed is better known as the *Minimal Naming Game* first introduced by Baronchelli *et al.* (2006a,b), hence we refer to it as the *Minimal NG Strategy*. The Minimal NG Strategy has its roots in opinion dynamics (Castellano *et al.*, 2009) and is related to the models of Fu & Wang (2008); Blythe (2009). For a detailed exposition of the most important aspects of the Minimal NG Strategy please see Baronchelli (2012).

The Minimal NG strategy brings together the Baseline NG Strategy and the Imitation Strategy. From the baseline strategy it keeps the ability to store multiple names per object with the difference that in case of a successful game (i.e. the listener knows the name), both agents remove all other competing names for that object as shown in Figure 2.9. Since this update cannot be seen as part of f_{invent} or f_{adopt} a fifth function $f_{align}(n, t, a)$ needs to be added to the strategy. Alignment takes place at the end of a language game, in this case, by both participating agents.

In the Minimal NG Strategy alignment depends on the communicative outcome of the game. At the end of a successful game $f_{align}(n, t, a)$ removes



(a) Update mechanics (b) Experimental results

Figure 2.8: (a) Update mechanics for the "stubborn" variant of the Imitation Strategy. The speaker utters the word "xuzuli" for the chosen topic. In case of failure the listener only adopts when he did not know a name yet. (b) Communicative and alignment success for the Strict Adoption Strategy. The strategy fails at reaching acceptable levels of communicative success.

all pairings (n, o) for which $o \neq t$. In successful games the Minimal NG Strategy behaves identical to the Imitation Strategy whereas in failed games it behaves like the Baseline NG Strategy.

In the Minimal NG Strategy communicative success and alignment success no longer overlap. It is possible for agents to communicate successfully about an object and not yet have aligned on a preferred name for that object. In Figure 2.10a alignment success starts out slower than communicative success but finally catches up and both reach full success. Compared to the Imitation Strategy alignment converges drastically faster as shown in Figure 2.10b. The Minimal NG Strategy exhibits a fast transition as discussed by Baronchelli *et al.* (2006b) where the Imitation Strategy does not.

So far we have presented three minimal strategies to complement the Baseline NG Strategy. Two strategies tried to reach alignment by limiting the agents to store multiple names per object. From these two only the adaptive agents of the Imitation Strategy were able to reach alignment and communicative success. The Minimal NG strategy showed faster alignment than the Imitation Strategy by combining properties of the Baseline NG Strategy and the Imitation Strategy.

The reason why the Minimal NG Strategy converges significantly faster than the Imitation Strategy is because it implements a somewhat extreme version of *lateral inhibition*. Agents inhibit all competing names by removing them. We now turn our attention to strategies implementing a more gradual inhibition scheme.

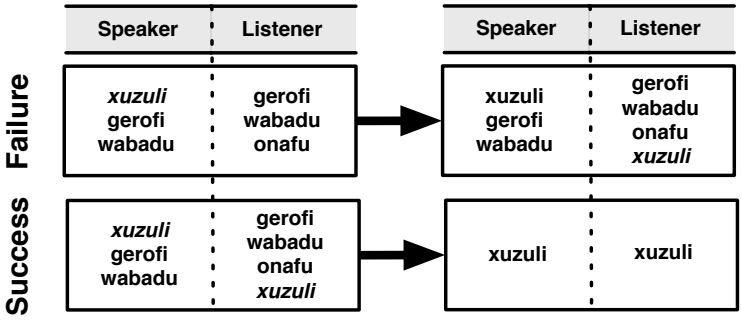
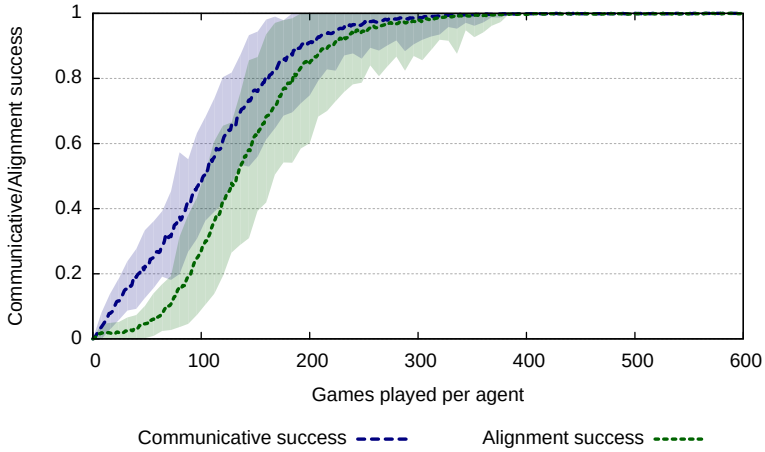


Figure 2.9: Update mechanics for the Minimal NG Strategy. The speaker utters the word "xuzuli" for the chosen topic. In case of failure the mechanics are the same as those form the Baseline NG Strategy (see Figure 2.2). In case of success they correspond to that of the Imitation Strategy (see Figure 2.5).

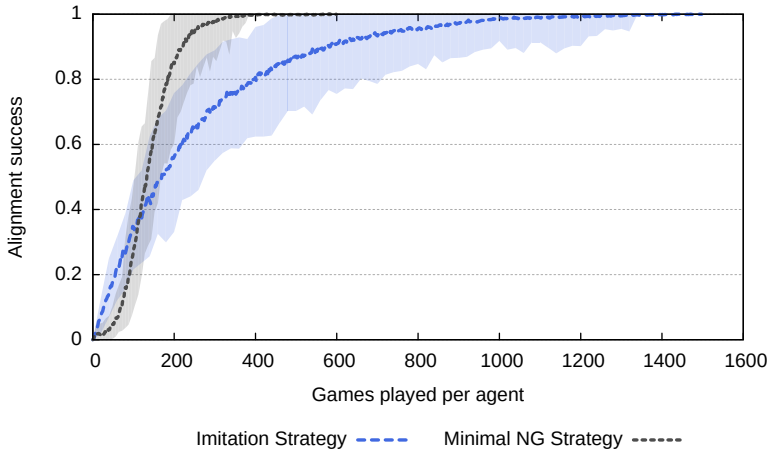
2.4 Lateral inhibition strategies

The term "lateral inhibition" comes from neurobiology and refers to the capacity of an excited neuron to reduce the activity of its neighbours (Hartline *et al.*, 1956). It is popular in connectionist modeling of pattern learning (Amari, 1977). As a mechanism in language games it was first proposed by Steels & Kaplan (1999) for the Talking Heads Experiment (Steels, 1999). In the context of the Naming Game it was introduced by Steels (2000). Lateral inhibition has since been widely used to dampen both form and meaning competition (Steels & Kaplan, 2002; Vogt, 2000; Vogt & Coumans, 2003)Before this, either ad hoc strategies or a form of frequency strategy (discussed in the next section) was used to tackle the problem of name competition. Keep in mind that the first minimal strategies were introduced only in 2005. Chronologically lateral inhibition strategies thus preceded the minimal strategies.

Lateral inhibition strategies aim to improve the previous strategies by extending the representational capabilities of the agents. Each pairing of a name with an object is scored by a real value in the interval $]0, 1]$. When multiple names compete for the same object identifier the score keeps track of the most popular or conventional name. Names are removed from the lexicon when their scores become equal to or lower than zero. Lateral inhibition refers to the process of agents actively inhibiting competing names at the end of each game. This inhibition is achieved, not by removing the competitors as



(a) Minimal NG Strategy



(b) Comparison with the Imitation Strategy

Figure 2.10: (a) Communicative success and alignment success for the Minimal NG Strategy. The agents reach full communicative success and alignment success. (b) Comparison of Imitation and Minimal NG Strategy in terms of alignment success. (a) and (b) population size: 50, object size: 5, number of simulations: 100, error: 5th to 95th percentile.

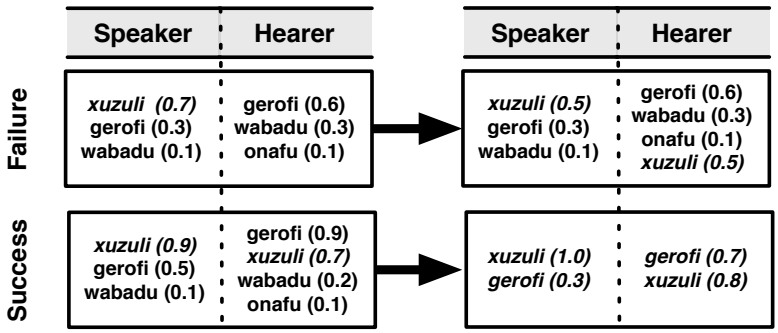


Figure 2.11: Update mechanics for the Basic Lateral Inhibition Strategy. The speaker utters the word "xuzuli" for the chosen topic. In case of failure the listener adopts the new name with $s_{initial}$ and the speaker decrements the spoken name with δ_{dec}. In case of success both agents increase the score of the spoken name with δ_{inc} and decrease (inhibit) the scores of the competitors with δ_{inh}. Here $s_{initial} = 0.5$, $\delta_{dec} = \delta_{inh} = 0.2$, and $\delta_{inc} = 0.1$.

in the Minimal NG Strategy, but by lowering their score.

Production $f_{produce}$ takes the score into account by preferring the name with the highest score when multiple possibilities are available. When the highest score is shared among multiple names a random name form this list is chosen. Invention and adoption initialize the score to $s_{initial}$. Interpretation remains the same as in the Baseline NG Strategy.

We present two lateral inhibition strategies which share the above changes but differ in their alignment f_{align}. The first is based on the proposal spelled out by Steels & Belpaeme (2005). This strategy, which we call the *Basic Lateral Inhibition Strategy*, introduces three parameters, δ_{inc}, δ_{inh}, and δ_{dec}, all influencing the scores of names during alignment. At the end of a failed game the speaker lowers the score of the used name by δ_{dec}, the listener adopts with the initial score $s_{initial}$. In case of communicative success both agents increase the score of the used name by δ_{inc} and decrease the score of all competitors by δ_{inh}. By default $s_{initial}$ is 0.5, scores are capped at 1.0 and names with $s \leq 0$ are removed. Figure 2.11 shows the update mechanics visually. Note that with parameters $s_{initial} = \delta_{inh} = 1$ and $\delta_{inc} = \delta_{dec} = 0$ the Basic Lateral Inhibition Strategy is identical to the Minimal NG Strategy.

The second lateral inhibition strategy differs only in that instead of incrementing or decrementing the score of a name with a given δ, it interpolates

towards either 1 or 0 as follows:

$$\text{reinforcement:} \qquad s \leftarrow s + \delta_{inc}(1 - s) \tag{2.4}$$

$$\text{inhibition:} \qquad s \leftarrow s - \delta_{inh} s \tag{2.5}$$

This update scheme for lateral inhibition was used for the Naming Game by Vogt & Coumans (2003) and later De Beule & De Vylder (2005) among others. It can be interpreted as the Rescorla-Wagner/Widrow-Hoff rule as described by Sutton & Barto (1981). A consequence of interpolations is that scores never reach 0 or 1 and agents thus never "forget" names, just like the Baseline NG Strategy. However, the agents develop preferences using the scores and should thus reach alignment success.

2.4.1 Experimental results for lateral inhibition strategies

Expectations are that the more subtle updating mechanics of the lateral inhibition strategies should result in faster alignment when compared to the Minimal NG Strategy. As explained the Minimal NG Strategy is a special case of the Basic LI Strategy which, compared to that strategy, limits the memory capacity of the agents by resorting to a complete removal of competitors instead of a gradual inhibition. The impact of lateral inhibition, however, turns out to be relatively small. Figure 2.12 shows alignment success for both lateral inhibition strategies and the Minimal NG Strategy (see caption for experimental parameters). Although convergence is faster for both in the beginning and only for the Basic LI Strategy in the end, the increase remains limited. This is all the more surprising since alignment, which is what has changed between these strategies, is doing the bulk of the convergence and not invention or adoption. In the same figure the vertical dashed line shows the point at which for all three strategies invention and adoption have come to a complete stop.

In the Interpolated Lateral Inhibition Strategy alignment slows down considerably, even to such an extent that the Minimal NG Strategy takes over. This can be understood by looking more closely at the update rules for both lateral inhibition strategies. Interpolation implies that the further away from an extreme the more the score will jump toward that point if possible. For example with $\delta_{inh} = \delta_{inc} = 0.3$ and $s_{initial} = 0.5$ it takes 11 exposures (i.e. increments) to reach a score $s > 0.99$ while it takes only 2 inhibitions to drop from that score of 0.99 to below 0.5 which is the initial score. Agents thus become ever more adaptive while many studies have shown, on the contrary, that frequency effects directly impact entrenchment (Bod *et al.*, 2000; Bybee, 1998).

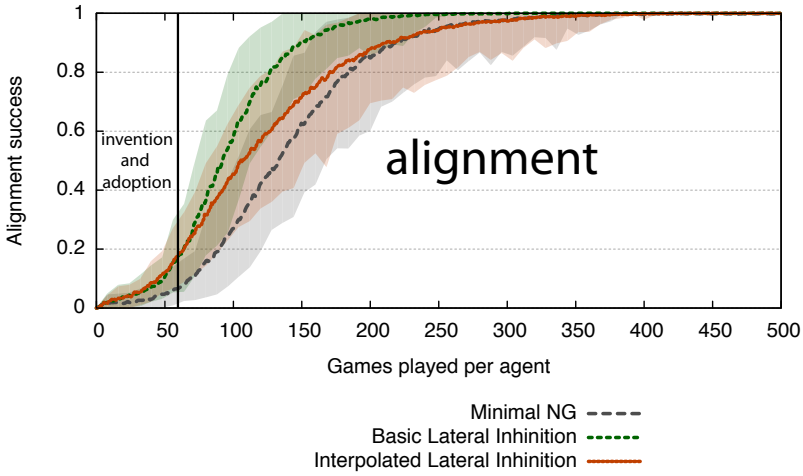
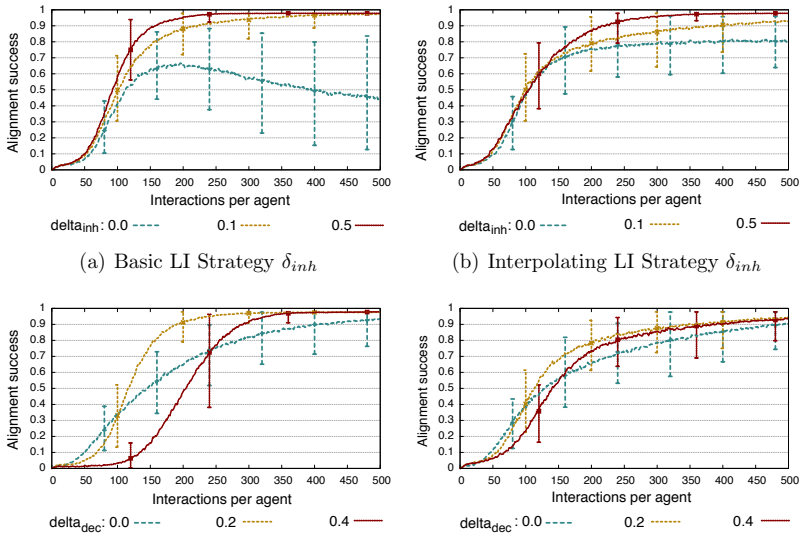


Figure 2.12: Alignment success for the lateral inhibition strategies. Parameters: population size: 50, object size: 5, total games: 10000, number of simulations: 100, error: 5 to 95 percentile. For both lateral inhibition strategies $\delta_{inc} = \delta_{dec} = 0.1$, $\delta_{inh} = 0.5$ and $s_{initial} = 0.5$.

The lateral inhibition strategies have parameters influencing their performance. The inhibition parameter δ_{inh} regulates the strength of the inhibition at the end of a successful game. The higher its value the more strongly competitors are inhibited. For example in the Basic Lateral Inhibition Strategy, with $\delta_{inh} = 0.5$ a name needs only be inhibited twice in a row to be certain of deletion and only once starting from $s_{initial} = 0.5$. On the other hand a value of 0.0 for δ_{inh} effectively disables inhibition and disallows the agents from reaching alignment success for both strategies. Results for different values of the inhibition parameter are shown in Figure 2.13(a-b) for both the Basic and Interpolating LI Strategy. The tendency shows that for both strategies a higher value improves alignment, although from the results for the Minimal NG Strategy we have learned that extreme values (e.g. $\delta_{inh} = 1.0$) again negatively impact alignment.

The story is different for δ_{dec} as shown in Figure 2.13(c-d). Disabling this functionality by setting $\delta_{dec} = 0.0$ is detrimental for alignment in the long run. As opposed to δ_{inh} a higher value for $delta_{dec}$ does not result in faster alignment. For example with $delta_{dec} = 0.4$ alignment takes significantly longer to reach its transition point.



(a) Basic LI Strategy δ_{inh}

(b) Interpolating LI Strategy δ_{inh}

(c) Basic LI Strategy δ_{dec}

(d) Interpolating LI Strategy δ_{dec}

Figure 2.13: Impact of δ_{inh} and δ_{dec} on the Basic and Interpolating Lateral Inhibition Strategy. Alignment success for the Basic (a) and Interpolating (b) Lateral Inhibition Strategy for different values of δ_{inh}. In both (a) and (b) $\delta_{dec} = \delta_{inc} = 0.1$ and δ_{inh} takes values 0.0, 0.1 and 0.5. (c and d) Alignment success for the Basic (c) and Interpolating (d) Lateral Inhibition Strategy for different values of δ_{dec}. In both (c) and (d) $\delta_{inc} = \delta_{inh} = 0.1$ and δ_{dec} takes values 0.0, 0.2 and 0.4. (a-d) population size: 50, object size: 5, total games: 10000, number of simulations: 100, error: 5 to 95 percentile.

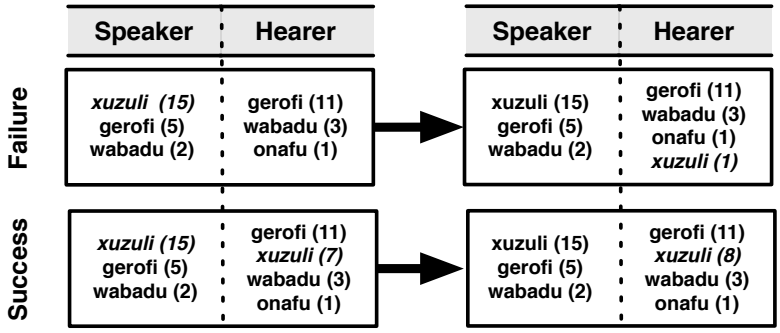


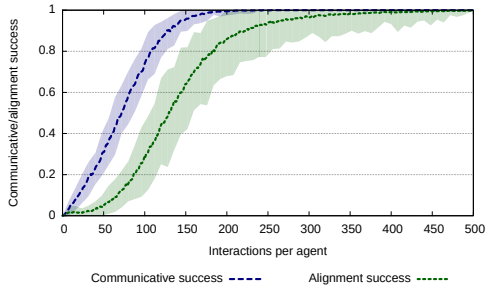
Figure 2.14: Update mechanics for the Frequency Strategy. Only the names for the topic object are shown. In case of failure the listener adopts the new name with frequency 1. In case of success the listener increments the frequency of the name with 1. Note that only the listener ever updates frequencies.

2.5 The Frequency Strategy

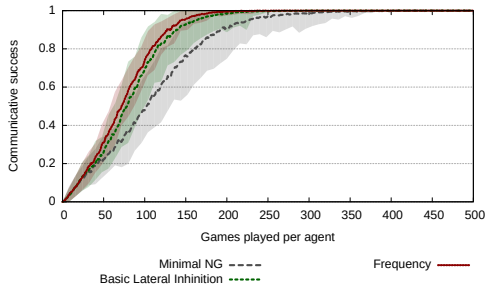
The Frequency Strategy was first proposed by Steels & McIntyre (1999). Instead of keeping a name-object pair $< n, o >$ agents following a frequency strategy keep a triplet $< n, o, f >$ with f an integer in the range $[1, \infty]$ representing the number of times that name was heard (not spoken).

Starting from the baseline strategy the production function $f_{produce}(o, a)$ is extended so that in case there are multiple names paired with o the one with the highest frequency is uttered. In the exceptional case that multiple names share the highest frequency a random one of those names is chosen. The learning functions f_{invent} and f_{adopt} differ only from the Baseline Strategy in that they initialize the frequency f to 1. Just like in the Minimal NG, the Frequency Strategy also needs an alignment function $f_{align}(n, t, a)$ which does nothing more than update the frequency of the spoken word, but only for the listener. The reason only the listener updates is that the frequency should capture what is the preference of the other agents and his own should just be that of the others. If the agent would also increase the frequency as speaker he would, half of the time, amplify his own preference rendering him less adaptive. De Vylder (2007)(p. 139) delves deeper into this issue and compares the impact of hearer-only versus hearer-and-speaker alignment in Naming Games. These updating mechanics are also illustrated in Figure 2.14.

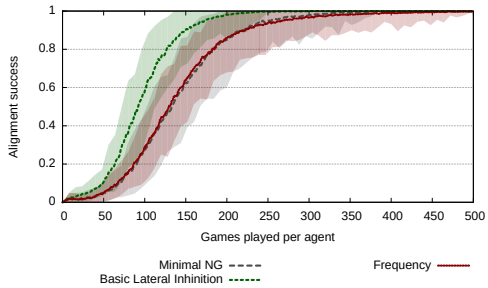
In the Frequency Strategy, just like in the Baseline NG Strategy, agents



(a) Frequency Strategy



(b) Communicative Success



(c) Alignment Success

Figure 2.15: Experimental results for the Frequency Strategy. (a) Communicative and alignment success for the Frequency Strategy. (b) Comparison of Minimal NG, Basic LI and Frequency Strategy in terms of communicative success. (c) Comparison of Minimal NG, Basic LI and Frequency Strategy in terms of alignment success. (a), (b) and (c) population size: 50, object size: 5, total games: 10000, number of simulations: 100, error: 5 to 95 percentile. The Basic LI Strategy in (c) used parameters $\delta_{inc} = \delta_{dec} = 0.1$, $\delta_{inh} = 0.5$ and $s_{initial} = 0.5$.

store and keep all names they ever encountered, which means they will also reach full communicative success as shown in Figure 2.15a. The population however reaches communicative success considerably faster because only the most frequent names will be used in production and so many invented names won't spread through the entire population. And indeed Figure 2.15b shows that communicative success with the Frequency Strategy much improves upon the Minimal NG Strategy, which itself improved on the Baseline NG Strategy. The reason for this lies in the fact that the Minimal NG Strategy deletes competing names in successful games, which might lead to a failure if one of the competing names is encountered later on. For alignment success we do not see the same gain as with communicative success. Alignment success remains below that of the Basic LI Strategy and almost coincides with that of the Minimal NG Strategy. Why alignment success between the Minimal NG Strategy and the Frequency Strategy is near identical remains an open question to me.

2.6 Conclusion

In this chapter I focused on the problem of form competition and the Naming Game that introduces it. The Naming Game is designed in such a way that it evokes only the conventionalization problem of name or word form competition. By equating meanings with holistic objects and perfect feedback the problem of meaning-related uncertainty is avoided. The only problem the agents need to solve is that of bootstrapping and maintaining a minimal set of names for a set of objects.

By presenting and comparing different strategies I have shown that multiple solutions exist to the problem of name or word form competition. Alignment, either by removing or inhibiting competitors or by counting frequencies, is crucial for a population to successfully play the Naming Game. All these successful strategies, except the Imitation Strategy, showed a fast transition from low levels of alignment to high levels of alignment. This sudden rise in the emergence of the communication system is also observed in the formation of basic human communication systems (Galantucci, 2005).

From the Imitation Strategy we learned that sometimes, less is more. By reducing the representational capabilities, the agents were capable of reaching full alignment because it gave rise to a winner takes all dynamics. Imitation agents can be seen as maximally adaptive, in that they immediately change their preference upon an exposure. Adaptivity, in the sense that agents need to change their name inventory based on their communicative encounters, is indeed a crucial ingredient for the emergence of a minimal naming system.

The limited representational capabilities of the Imitation agents came at

Strategy	Baseline	Imitation	Minimal NG	LI	Freq
(1) full CS	yes	yes	yes	yes	yes
(2) WFC solved	no	yes	yes	yes	yes
(3) scoring	no	no	no	yes	yes
(4) fast transition	N/A	no	yes	yes	yes
(5) competition damping	no	no	yes	yes	no

Table 2.1: Overview of Naming Game strategies. Depicts whether (1) strategies could reach full communicative success (CS), (2) could solve word form competition (WFC), (3) implemented a scoring mechanism, (4) showed a fast transition in alignment, and whether (5) they implemented an explicit damping of competitors.

the cost of slow alignment. The Minimal NG Strategy extends the abilities of agents in terms of both representation and processing. On the representational level it allows agents to maintain lists of competing names, which in turn allows the agents to be slightly less "adaptive" when confronted with a novel name. Agents can store the name as a competitor of the other names he has heard for that object. Only when a name is either heard os used successfully again will he remove its competitors. This change lead to considerably faster convergence both in terms of communicative as alignment success.

The Minimal NG Strategy is a special case of a more general type of strategies that rely on lateral inhibition. Successful words are awarded and competing words are damped. Such a scheme was implemented in a gradual way by adding a score to each name object pairing. With this added representational capability agents did not need to remove competing names so hastily as in the Minimal NG Strategy. Instead, based mainly on the inhibition parameter δ_{inh}, agents have a more gradual memory. It turned out that lower values for inhibition did not improve alignment success.

From these results it seems that adding a scoring mechanism yields only marginal improvements in terms of communicative and alignment success. Although this holds true for the Naming Game we will see in the following chapter that when other problems are introduced next to the problem of word form competition, slower update mechanics become increasingly beneficial and even crucial for alignment.

The last discussed strategy never removes or inhibits names and thus differs considerably from all previous extensions to the Baseline Strategy. It extends the baseline by adding the representational capability of keeping a frequency for each name object pair. This frequency only needs to be updated as a listener and can be done regardless of communicative success. In production a speaker following the Frequency Strategy chooses the most

frequent corresponding name. This strategy reaches full communicative and alignment success. While it does not improve alignment success compared to the Minimal NG Strategy it does improve communicative success. Because it does not implement a direct competition inhibition like the lateral inhibition strategies, alignment moves slightly slower. An overview of all strategies is given in Table 2.1.

Part II

Uncertainty in lexical systems: competitive strategies

> Point to a piece of paper. And now to its shape - now to its color - now to its number....How did you do it?
>
> Wittgenstein, Philosophical Investigations

Introduction

Many words in natural language refer to categories and not to objects as a whole. Words such as "red" or "large" do not refer to objects, they rather refer to categories that belong to objects. As it is not possible to point to these properties the problem of *uncertainty* is introduced in word learning.

Quine (1960) famously described a similar problem through an example picturing an anthropologist studying the – unknown to him – language of a tribe. One of the natives utters the word "gavagai" after seeing a rabbit. How can, even after repeated uses of this word, the anthropologist come to know the meaning of "gavagai"? It could mean rabbit, an undetached rabbit part, food, running animal or even that it's going to rain. However, when we learn a language we do acquire the meanings of words with enough accuracy to be used in successful communication. It is an ongoing research question as to how we achieve this task.

From the age of around eighteen months to the age of six years, children acquire on average nine new words a day (or almost one per waking hour) (Bloom, 2000). In her research on word meaning acquisition, Carey (1978) introduced the distinction between fast mapping and slow mapping. Fast mapping is the process in which the first operational, yet partial, mapping is made between the domain of form and that of meaning. It is called fast because it involves only a few exposures to a novel word, yet it creates only a fragile first hypothesis [see (Horst & Samuelson, 2008; Vikram K. & Markman, 2001; Wilkinson & Mazzitelli, 2003) for more on fast mapping]. Slow mapping is the process, which can take up to several years, of further refining this fragile entry to its full meaning (Carey & Bartlett, 1978; Horst & Samuelson, 2008).

A large body of research has focused on the problem of creating the first initial mapping or has focused on meanings for which the concept is already known and all that remains to be done in learning is correctly mapping a word to that pre-established concept (Bloom, 2000). The implicit assumption is that learners have access to a number of potential meanings and need to choose (or guess) the correct one. Under these assumptions the problem of

word learning is seen as a mapping problem and the uncertainty that learners face is one of *mapping uncertainty*. Within the mapping approach, several solutions to the problem of mapping uncertainty have been theorized. One proposal is that the learner is endowed with several word learning constraints (also called biases or heuristics) that guide him towards the right mapping [see for example (Gleitman, 1990; Markman, 1992; Landau *et al.*, 1988), Smith (2005a) gives a good overview in the introduction].

As Smith *et al.* (2009) suggests, such heuristics can aid in reducing mapping uncertainty but not completely remove it. A learning scheme, called *cross-situational learning*, has been proposed by various authors (Pinker, 1989; Siskind, 1996; Gillette *et al.*, 1999) to nevertheless learn the correct word-meaning mappings under uncertain conditions. The idea behind cross-situational learning is that each exposure to a word presents the learner with a set of possible meanings, and thus possible mappings. Given multiple exposures the learner should be able to eliminate incorrect mappings by comparing these sets. One example is that the learner could assume the intended meaning is the intersection of all sets.

In chapter 3 we investigate the problem of mapping uncertainty, where the meanings or concepts themselves are already established. Mapping uncertainty can be computationally investigated through a language game very similar to the Naming Game with the difference that agents do not know the intended meaning at first exposure and thus need to rely on cross-situational statistics. Just like in the Naming Game only a single word is spoken per game and this word is assumed to refer to either a category or one of the objects. These language games are called Minimal Guessing Games.

Many words, however, express multiple categories or incorporate many different semantic domains and have a complex structure. Moreover almost all utterances are compositional: more than one word is used to cover the set of categories that the speaker intends to convey. Human languages are therefore compositional (words express complex meanings that can be combined into a compositional utterance) rather than holistic (a single word expressed all of the meaning). Agents now need to establish the meanings themselves, which in the reported experiments are sets of categories, and agree on their form. The uncertainty is no longer one of mapping but relates to the specifics of the meanings themselves, which I call meaning uncertainty. This much more difficult task is embodied in the Compositional Guessing Game.

Chapter 4 investigates whether the competitive cross-situational approach can be straightforwardly applied to the problem of compositional meaning uncertainty. The most common implementation of cross-situational learning relies on enumeration and subsequent reduction of possible meanings. When meanings themselves are to be constructed the hypothesis space grows

exponentially and enumeration becomes problematic. I call this standard implementation of cross-situational learning “competitive”, because it relies on a competition dynamics between competing mappings.

Chapter 3

Cross-situational strategies in Minimal Guessing Games

Computer models addressing the problem of meaning uncertainty started with the work of Hovland & Hunt (1961) on concept learning although it took another decade for Simon & Lea (1974) to present the problem as a search problem through a hypothesis space. During the past decade attention shifted from the problem of concept formation to that of word learning (Oates, 2003). Meaning uncertainty in word learning has been investigated through evolutionary language games from its very start. Language games introducing uncertainty about meaning have been called *Guessing Games* (Steels, 2001) indicating that the agents have to guess the meaning of novel words. In fact, in retrospect, one of the very first operationalized language games was a rather complex Guessing Game (Steels, 1995), although at that point the term Guessing Game was not yet introduced.

During the first decade of investigations in the language game paradigm, language games were generally given a name based on the *interaction script*. For example Vogt (2000) distinguishes between an observational game, a guessing game and a selfish game based on the script the agents follow. Steels & Belpaeme (2005) also defines a Guessing Game based on the interaction script. Sometimes even a particular language strategy (e.g. learning) is also tied to the definition of the language game.

In this thesis I try to differentiate language games based on the underlying *conventionalization problem* they introduce. For example in the previous chapter games that only invoke the problem of word-form competition without uncertainty about word meaning are called Naming Games irrespective of the interaction script of the game and irrespective of the particular strategy as indeed multiple strategies have been presented for the same game. Obviously you do need to provide a script in order to operationalize the experiment but

this is not the crucial defining ingredient of the language game. For example I will provide two scripts or interpretations of the language game introduced in this chapter.

The Naming Game discussed in the previous chapter imposed a series of restrictions in order to only focus on the problem of word form competition (see page 34). Most prominently it avoids the problem of uncertainty by imposing the following two constraints:

3. Objects do not show any internal complexity and names thus can only refer to objects as a whole and not to categories or features.

4. At the end of the game both participants know the intended referent and can thus unambiguously make a pairing between name and object.

In this chapter we lift the fourth constraint so that agents do not know, at first exposure, the exact mapping between form and meaning. This introduces a limited variant of word meaning uncertainty which I call *mapping uncertainty*. Agents have to find out which words map to which individual referent or category. The term meaning uncertainty is reserved for cases where meaning exhibits internal structure and needs to be established within the population, a problem only introduced in the following chapter.

Language games that only introduces mapping uncertainty will be called *Minimal Guessing Games* as they represent the most minimal extension from the Naming Game. Note that the problem of word form competition is also inherited from the Naming Game. The population is thus faced with both the problem of mapping uncertainty (which word maps to which meaning) and word form competition (which will become the dominant word form to express a certain meaning). See Figure 3.1 for a schematic overview.

In order to solve the problem of mapping uncertainty, a learning method called *cross-situational learning*, has been proposed (Pinker, 1989; Siskind, 1996; De Beule *et al.*, 2006; Smith *et al.*, 2006; Vogt & Smith, 2005). Each usage event supplies the learner with a set of compatible meanings or mappings. The intended meaning must then either lie at the intersection of these possibilities or should be the one that is most often compatible.

Some other language game implementations have also removed the third constraint from the Naming Game and operate in a compositional or even continuous meaning space. For example in the Talking Heads experiment objects are represented as sets of continuous feature channels (Steels, 1999; Steels & Kaplan, 1999). In the Talking Heads and likewise in (Vogt, 2000; Smith, 2003b, 2005a) the continuous space is discretized using a discrimination component that makes use of discrimination trees. Nodes in the discrimination trees correspond to categories and meanings are taken to be a single category.

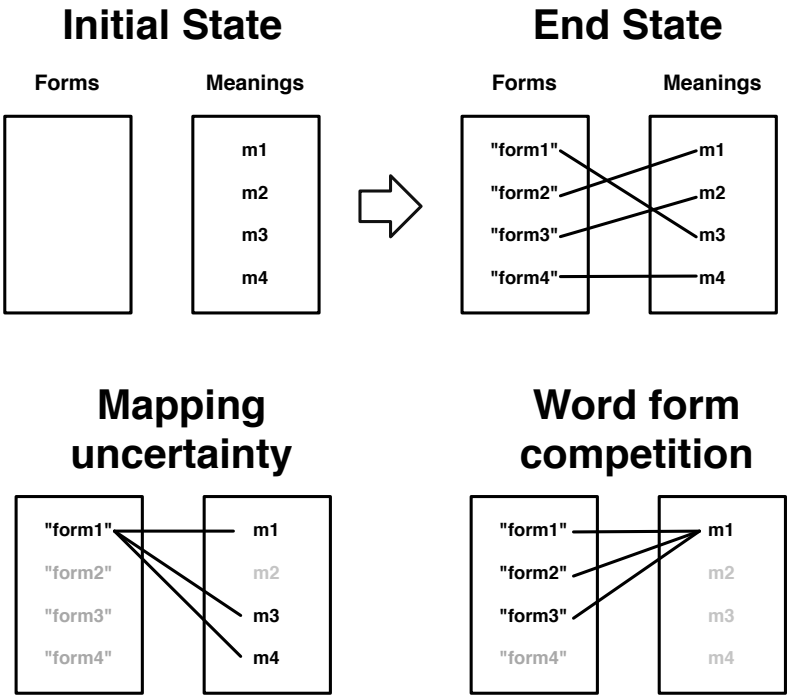


Figure 3.1: Lexical System and conventionalization problems in the Minimal Guessing Game. In the Minimal Guessing Game the agents need to establish a shared mapping of forms to meanings. The meanings are assumed to be atomic and are already given. Two conventionalization problems need to be solved. The problem of word form competition, inherited from the Naming Game and the new problem of mapping uncertainty.

In (Vogt & Coumans, 2003; Divina & Vogt, 2006; Vogt & Divina, 2007) objects are represented as sets of categories but again word meanings are assumed to express only a single category. As such all the above games fit best, although not perfectly, within the category of Minimal Guessing Games.

Experimental studies with human subjects have shown that both children and adults are capable of cross-situational learning (Gillette *et al.*, 1999; Piccin & Waxman, 2007; Smith & Yu, 2008). Smith *et al.* (2009) have however eloquently shown that this capability should not be overestimated. In these experiments subjects are generally confronted with multiple exposures to words and "objects" without telling the subjects how they map onto each other. Depending on the condition they are given either two, three or four words and objects at a time. Obviously the more words and objects are presented at an exposure the greater the uncertainty. In a second phase the participants are asked to point to one of the objects while presented with one of the words from the first phase. If the participants are successful it is assumed they must have stored some sort of cross-situational statistics during the first phase of the experiment.

Smith *et al.* (2009) compared the performance of the participants with that of a computational agent that did *not* employ cross-situational statistics. Instead of taking into account all exposures from the first phase this agent only looked at a single exposure when making a decision in the second phase. It turned out that in the 4 words, 4 objects condition this computational agent outperformed the human subjects. The reason was that in the test phase also only 4 objects (one of them the correct referent) were presented at each exposure. The non-cross-situational learner thus pointed to a random object in the intersection between the (limited) test set and the single exposure it remembered. The larger the total set of objects the higher the chance this intersection contains only the intended referent. Smith *et al.* (2009) have then performed new experiments in which during the test phase all objects were present during each test exposure. In these experiments the human subjects failed to perform significantly better in the 4 words, 4 objects case than the non-cross-situational learner.

3.1 Definition and script of the Minimal Guessing Game

The Minimal Guessing Game is the most minimal extension to the Naming Game that brings out the problem of mapping uncertainty. Meanings of words in the Guessing Game have in previous publications been presented in terms of referents and categories. Both flavours, in their Minimal variant,

can be summarised to have the following features:

1. A Minimal Guessing Game involves two participants, a listener and a speaker.

2. In a Minimal Guessing Game a speaker utters a single word to refer to (a) one object in a context or (b) one category of an object.

3. **(a)** Objects do not show any internal complexity and words thus can only refer to objects as a whole and not to categories or features.

   (b) The object shows internal complexity, implemented as a set of categories, and a word refers to a single category.

4. At the end of the game only the speaker knows the intended (a) referent or (b) category.

   (a) The listener only knows that the spoken word refers to one of the objects in the context.

   (b) The listener only knows that the spoken word refers to one of the categories of the object.

5. Uncertainty about the correct mapping for the referent/category of a word is the core problem of the Minimal Guessing Game.

6. As in the Naming Game, competing possibilities for the word of a referent/category (word form competition) are introduced and increase with the size of the population.

Although path (a) or (b) in the summary above are quite different with respect to how objects are represented they evoke the exact same conventionalization problem. In this thesis they are considered the same type of game (i.e. a Minimal Guessing Game) since both evoke the problem of lexical uncertainty in its most minimal way, namely as a mapping problem. Word meanings are assumed to be atomic and only a single word is uttered per interaction.

Following path (a) the script of a Minimal Guessing Game is that of the Naming Game from the previous chapter except for the feedback at the end as shown in Figure 3.2. Following path (b) results in another script that would require feedback from the speaker (i.e. pointing to the topic object). In the remainder of this chapter the script depicted in Figure 3.2 is followed.

More formally the Minimal Guessing Game discussed in this chapter requires objects or meanings $o_i \in O$ and agents $a_1 \cdots a_n \in P$ (population). For every game a context of objects is generated and two agents a_i and a_j

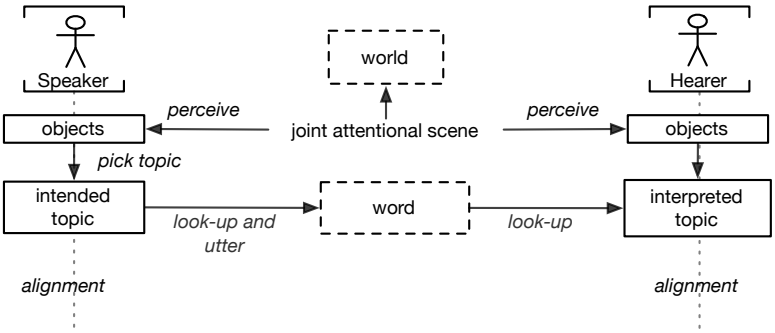


Figure 3.2: Possible script of a Minimal Guessing Game. Squared boxes represent data entities. Italicised text refers to processes. The look-up processes can fail and therefore have diagnostics and repair strategies associated with them. A language game experiment consists of thousands of these interactions.

are chosen from the population. One agent is assigned the role of speaker, the other of listener. Both agents are confronted with the same context of n objects. When these conditions are met the game can start.

1. The speaker mentally picks one object from the context (called the *topic*) and utters a word for it.

2. The listener interprets the word in the given context.

3. At the end of the game the listener does *not* receive perfect feedback like in the Naming Game but instead remains uncertain about the intended meaning of the speaker. Both agents have the opportunity to adapt their inventory of words.

To summarize, in this chapter we investigate the problem of establishing word to meaning mappings, almost identical to the Naming Game, but with the added problem that upon exposure to a word its meaning cannot be established with certainty.

3.2 Measuring success in the Minimal Guessing Game

The Minimal Guessing Game evokes two layers of competition. Mapping uncertainty leads the agents to enumerate multiple competing meanings for one particular word form. As long as this list of competitors is not reduced to one the agents remain uncertain about the conventionalized meaning. At the same time the problem of word form competition, inherited from the Naming Game, is still in play. Multiple words can compete to express the same meaning. Although this competition is also introduced from the very start, it can only be addressed when the mapping uncertainty starts to ease out. Indeed, how can two words compete for the same meaning when the meaning of the words in question is not yet clear. The Minimal Guessing Game is only successful when each word has only one meaning associated to it and each meaning only one word. The desired end state of the Minimal Guessing Game is thus not different from that of a Naming Game (see Figure 3.1).

Due to the lack of feedback, communicative success cannot be measured as was possible in the Naming Game. We can measure whether both agents associate the same preferred meaning to the spoken word but this information is not available to the agents themselves. Figure 3.1 depicts the state toward which the population should evolve, which is a single mapping for each expressible meaning. The figure also illustrates the two types of intermediary problematic states that need to be resolved. A measure for the Minimal Guessing Game should thus measure to what extent the population has reached the desired end state.

The usage-based measure *Mapping alignment* records at the end of every game (i) whether both listener and speaker prefer the same meaning for the spoken word (no mapping uncertainty for that word) and (ii) whether for that meaning the spoken word is their preference (no word form competition for that meaning). It is a Boolean measure which only returns 1 when speaker and listener share the exact same preferred mapping between form and meaning for the spoken word. This measure can be seen as a local variant of the measure called *Return* introduced by De Beule *et al.* (2006).

Mapping alignment is the combination of two measures, *Form alignment* and *Meaning alignment*. Form alignment corresponds to preferring the same form for a given meaning and meaning alignment to preferring the same meaning for a given form. Form alignment is thus the same as what Vogt & Coumans (2003) calls coherence. Note that these measures, just like mapping alignment, are external to the agents since the listener is never aware of the topic and the speaker does not know about preferences of the listener.

A successful strategy for the Minimal Guessing Game should arrive at full mapping alignment. Low form alignment signals that the agents have a problem with damping competing forms for the same meaning. Low meaning alignment means the agents have trouble arriving at figuring out the intended meanings for words. In this thesis I do not use the linguistic terms synonymy or homonymy as they denote much more complex processes and constitute different phenomena than the problems encountered in a Minimal Guessing Game.

A final measure which every strategy is tested on is *Acquisitions robustness*. This measure differs from all the above in that it assumes an acquisition setup with a learner agent and a tutor agent instead of a multi-agent "bootstrapping" setup. This measure mimics in computational simulation, a setup like those found in the experiments reported in (Gillette *et al.*, 1999; Piccin & Waxman, 2007; Smith & Yu, 2008). A single learner agent needs to learn a single novel word w for a target meaning m. To learn the target meaning the agent is presented a series of contexts in combination with the word w. The parameter $error_{input}$ determines the percentage of contexts in which the target meaning is *not* present. If $error_{input} = 0.0$ then all contexts contain m, if it is 0.1 then 10% of all contexts do not contain the target meaning m and thus represent an inconsistent input. *Acquisition robustness* measures for a series of exposures (or games) the percentage of times the agent "correctly" associated m with w. If acquisition robustness remains high under high levels of $error_{input}$ it means the strategy is robust against inconsistent inputs.

3.3 A baseline non-cross-situational strategy

A language strategy needs to supply both representational and processing aspects. We start with a strategy that builds on the Minimal NG strategy discussed in the previous chapter. We will call this strategy the Baseline GG Strategy.

In this strategy agents maintain a bidirectional memory of word-meaning mappings (w, o), where the meaning is a reference to an object o. The agents *cannot* represent uncertainty about whether this mapping is correct. Not by keeping a certainty score, neither by maintaining lists of possible candidates. Their representational power is thus the same as that of agents following the Minimal NG Strategy.

For production the Baseline GG Strategy supplies $f_{produce}(o, a)$ which looks up object o in the memory of agent a and returns the associated word. If multiple words for o exist it returns a random one, if no word is found $f_{invent}(o, a)$ creates a new word w and adds the association (w, o) to the inventory of the agent. Both $f_{produce}$ and f_{invent} are the same as for the

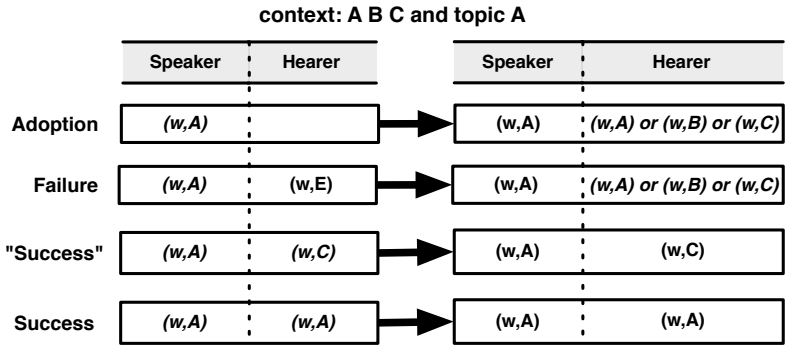


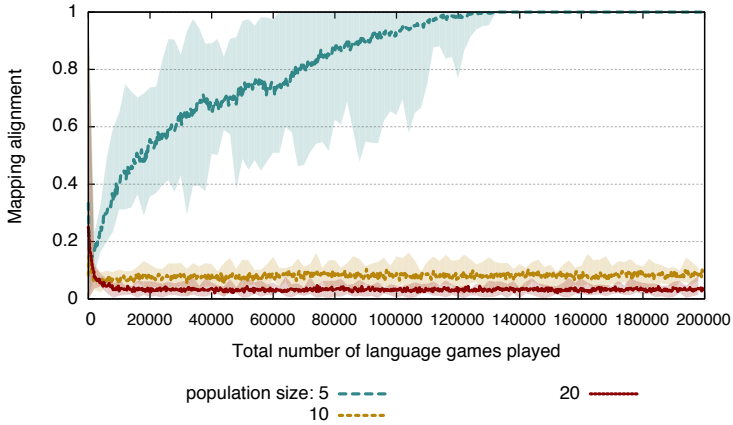
Figure 3.3: Update mechanics for the meaning of a word in the Baseline GG Strategy. Given context [A,B,C] and topic A (only known to speaker), the speaker utters the word (w,A). In case of adoption the listener does not know w. In case of failure the agent knows w but has an associated meaning not present in the context. In the third case, the listener associates a different meaning to w but this meaning is also present in the context. The agent has no means to find out whether it is correct and does not update. Finally there is the case in which their meanings are the same.

Minimal NG Strategy.

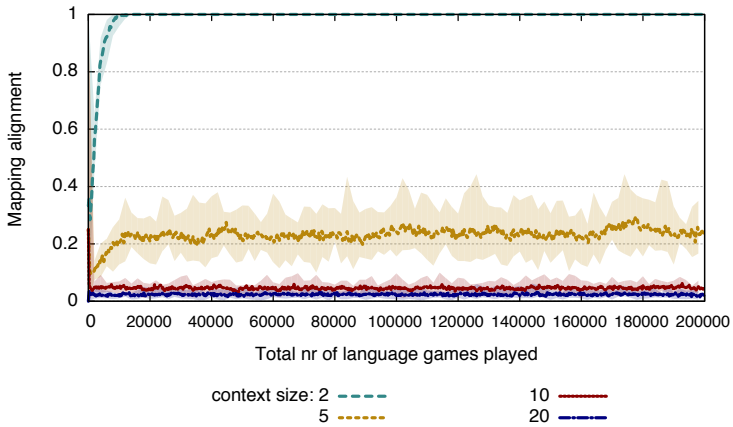
In interpretation the listener looks up the spoken word w and retrieves the associated meaning or object o. Because of the representational restrictions there will never be more than one entry for a given word. There are three possible scenarios in interpretation for the Baseline GG Strategy:

1. The listener does not know the spoken word w.
2. The listener knows the word w and the associated meaning does *not* match one of the objects in the context.
3. The listener knows the word w and the associated meaning does match one of the objects in the context.

In the first case the listener adopts the novel word. Function $f_{adopt}(w, C, a)$ takes the spoken word w, the agent a and the context C and adds association (w, o) with $o \in C$ randomly chosen from C to the memory of agent a. In contrast to the Naming Game this association can be incorrect. The chance for the "guess" to be correct is $\frac{1}{|C|}$. The second case signals that the agent may have made an incorrect guess when last adopting. He removes the word



(a) Scaling population size



(b) Scaling context size

Figure 3.4: Scaling in the Baseline GG Strategy in combination with the Minimal NG Strategy. Only in the most limiting scenario's can agents reach mapping alignment and even still at a very slow pace. (a) total number of objects: 50, context size: 4, total games: 200000, error: min/max, number of simulations: 12, population sizes: 5, 10 and 20. (b) total number of objects: 50, population size: 5, total games: 200000, error: min/max, number of simulations: 12, context sizes: 2, 5, 10 and 20.

and makes a new guess by adopting the word again. In the last case the listener has no reason to change the association. Note however that this does not mean both agents associated the same meaning. All that is certain is that both associate an object found in the context. The different scenarios schematically depicted in Figure 3.3.

The mechanics just sketched and schematically depicted in Figure 3.3 are *non*-cross-situational in that agents do not keep or use cross-situational statistics. Instead agents keep taking guesses as long as their hypothesis is invalidated.

When agents achieve alignment on the same meanings for the same words (solved mapping uncertainty) they still have different words referring to the same meaning (word form competition). The strategies for coping with mapping uncertainty thus need to be complemented with a strategy for form competition from the previous chapter.

The Baseline GG Strategy is closest to the Minimal NG Strategy so it is only natural to add the damping mechanics of that strategy (details can be found at page 48, Chapter 2, Section 2.3). Since there is no measure of communicative success the Minimal NG alignment is run in every game, although only for the listener. In short, the listener removes all words that associate the same preferred meaning as that of the spoken word.

Agents following the Baseline GG Strategy combined with the Minimal NG Strategy can reach full mapping alignment as long as the number of agents and the context is kept small. Figure 3.4(a) shows that increasing the size of the population from 5 to 20 has a drastic impact on speed of alignment. In fact even for a population of only 10 agents there is no significant sign of improvement over the first 200000 games. For a population of 10 agents each agent on average participated in 40000 games of which 20000 as listener. For a total of 50 objects each agent thus had on average 400 update opportunities for each object. The graphs (esp. for a population of 10 and 20) show a decrease in the beginning. This is simply an effect that the more competitors (both form and meaning) are going around in the population the higher the chance for a low average mapping alignment. In the beginning there are on average less competitors since they are still spreading and thus on average Mapping alignment is slightly higher.

Figure 3.4(b) shows the impact of increasing context size, which is a key parameter since it most influences the amount of uncertainty upon hearing a novel word. Here the context was increased from 2 to 20 for a total number of objects of 40. The population size was kept constant at 5. These results show, but not yet explain, that the Baseline GG Strategy has a problem with scaling both in terms of population size and context size.

As a first hypothesis we investigate the role of the strategy for damping

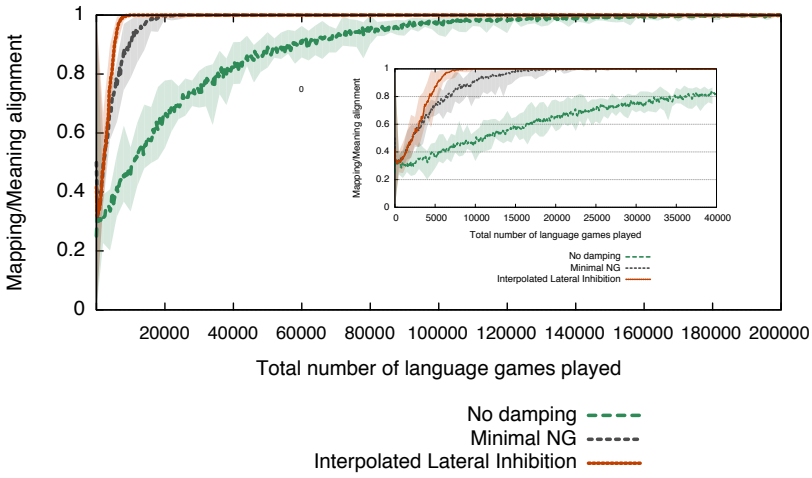


Figure 3.5: Meaning and mapping alignment with and without damping of word form competition. Meaning alignment can be reached for small populations and small context size even without a damping mechanism, albeit very slowly. A Naming Game strategy improves mapping alignment drastically. More subtle updating mechanics such as the interpolating Lateral Inhibition improve this even further. The embedded graph zooms in on the first 40000 games in order to better compare the bootstrapping behaviour of the Minimal NG and interpolated LI conditions. Parameters: population size: 5, total number of objects: 40, context size: 3, error: min/max, number of simulations: 12.

competing word forms. This can be achieved by calling the f_{align} function from different Naming Game strategies at the end of the language game. The only notion this function relies on is that of a word form competitor (words referring to the same object). We measured mapping alignment[1] for a population of 5 agents, a total number of 40 objects and a context size of 3 for the following three conditions:

- Absence of word form competition (no Naming Game updates)
- Minimal NG updates as were used for the experiments reported in Figure 3.4.
- The interpolated lateral inhibition update mechanics as explained on page 51, Section 2.4, Chapter 2. $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$.

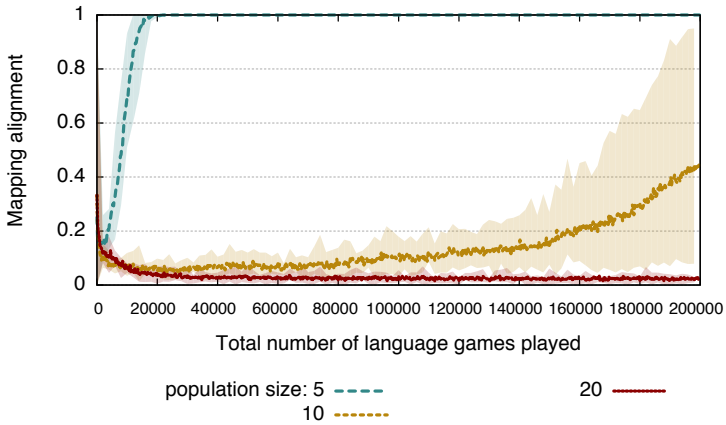
Interestingly Figure 3.5 illustrates that the lack of an associated Naming Game Strategy heavily impacts on meaning alignment. Remember that the NG Strategy only dampens word form competition and does not align meanings of words. What this shows is that eliminating competing forms also indirectly helps in aligning the correct meanings. The graph also shows that a more subtle updating mechanism like the interpolated Lateral Inhibition scheme leads to faster mapping alignment. This is again explained by the interplay between aligning both form and meaning. It is better not to eliminate form competitors too fast, since first meanings need to become aligned. The more subtle interpolated Lateral Inhibition Strategy gives the meanings time to align (through the cross-situational dynamics) and only at a much slower pace takes care of word form competition.

Running the scaling experiment from Figure 3.4 again, but this time with the Minimal NG Strategy replaced by the interpolated Lateral Inhibition Strategy, shows that scaling has improved significantly, but still remains poor (see Figure 3.6). This means that the Baseline GG Strategy itself has an issue with scaling.

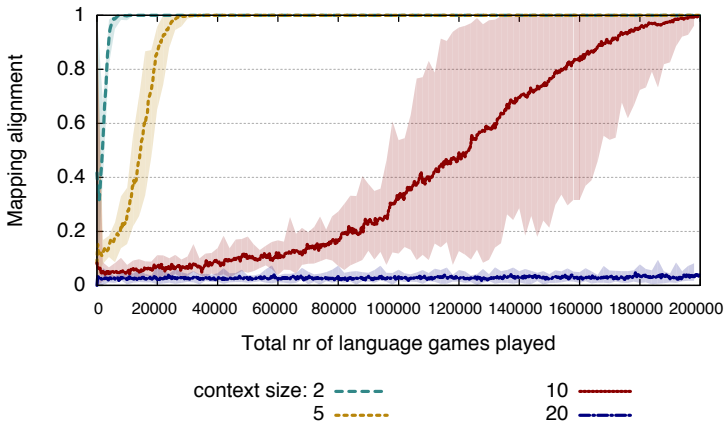
By looking closer at a worked out example we can see that the chance to align on a meaning is indeed very small. Given a context C and the set of all objects O and a single word w for which agent a_{good} has the correct association $< w, o_1 >$ and agent a_{bad} has another association $< w, o_2 >$. A language game in which a_{good} and a_{bad} interact using word w has three possible outcomes:

(1) positive: a_{bad} learns the correct association $< w, o_1 >$.

[1]For the setup without an associated Naming Game strategy only meaning alignment is measured since form alignment cannot be reached without a NG strategy. As a consequence also mapping alignment would be very low.



(a) Scaling population size



(b) Scaling context size

Figure 3.6: Scaling of Baseline GG Strategy in combination with the interpolated Lateral Inhibition Strategy. Scaling is improved but still remains poor. (a + b) number of objects: 50, error: min/max, number of simulations: 12 $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$ (a) total context size: 4, population sizes: 5, 10 and 20. (b) population size: 5, context sizes: 2, 5, 10 and 20.

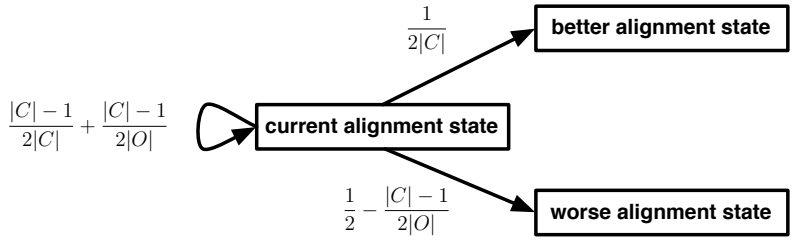
(2) neutral: There is no fundamental state change, in that a_{bad} still has an incorrect association and a_{good} keeps the correct one.

(3) negative: a_{good} forgets the correct association and learns a faulty one.

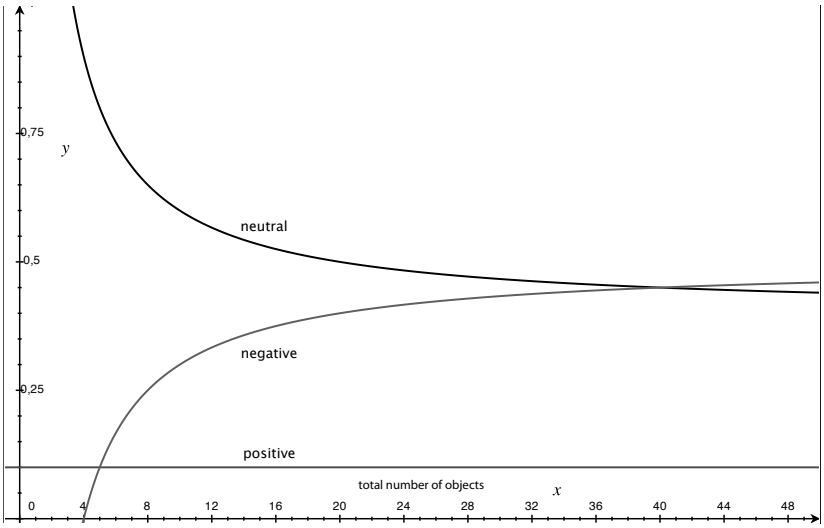
To arrive in state (1), a_{bad} is required to be listener (chance $\frac{1}{2}$) *and*, in adoption, he has to pick the correct association (chance $\frac{1}{|C|}$). So only with a chance of $\frac{1}{2|C|}$ will a positive transition occur. State (2) is reached when a_{bad} is listener and picks an incorrect new association with a total chance of $\frac{|C|-1}{2|C|}$. There is also a small chance in which a_{bad} is speaker and $o_1 \in C$ in which case a_{good} will not update his association (chance $\frac{|C|-1}{2|O|}$). State (3), which is a worsening of the situation occurs when a_{bad} is speaker and o_1 is not part of the context (chance $\frac{1}{2} - \frac{|C|-1}{2|O|}$). Please refer to Figure 3.7 for a schematic representation. For a context of 4 objects and a total number of 40 objects the chance for an improvement is $\frac{1}{8} = 0.125$, the chance to remain in the same situation is 0.4125 and the chance to end up in a worse situation is 0.4625. Although the chance for improvement is small it does not render the goal of full mapping alignment impossible because once all agents do agree on the same meaning for a form this state is stable. To reach that stable state the population just requires a large sequence of favourable encounters, which is rare but not impossible. The larger the population size the larger this sequence becomes which explains the results shown in Figure 3.7a. The larger the context size the smaller the chance for such a favourable encounter becomes which explains the results shown in Figure 3.7b.

To understand these poor results in scaling let us look at the strategy's robustness against inconsistent input, which is an important feature when the population needs to *bootstrap* the lexicon. Inconsistency is the rule and not the exception in the initial phase of invention and adoption. As explained earlier Acquisition robustness measures the percentage of games in which an agent can maintain the correct target meaning m for a word w under different amounts of inconsistency.

As shown in Figure 3.8 even 1 inconsistent exposure per 100 leads to an incorrect association 15% of the time. Five percent of inconsistent exposures reduces the percentage of correct associations to on average sixty. The reason this strategy is so fragile to inconsistent input is that every inconsistent input makes the agent change the meaning to a random one from the set (which is guaranteed to be incorrect). Only when he randomly picks the correct association again can the agent maintain the correct association, at least until the next incorrect exposure.



(a)



(b)

Figure 3.7: Alignment chances and example for baseline GG Strategy. (a) Given a context C, the set of all objects O, the schema shows the chances that in one game, which offers an opportunity for alignment, the agents move to a better or worse state of alignment. See text for details and an example. (b) Given that $|C| = 5$ the chances are plotted (y-axis) for an increasing size of object O.

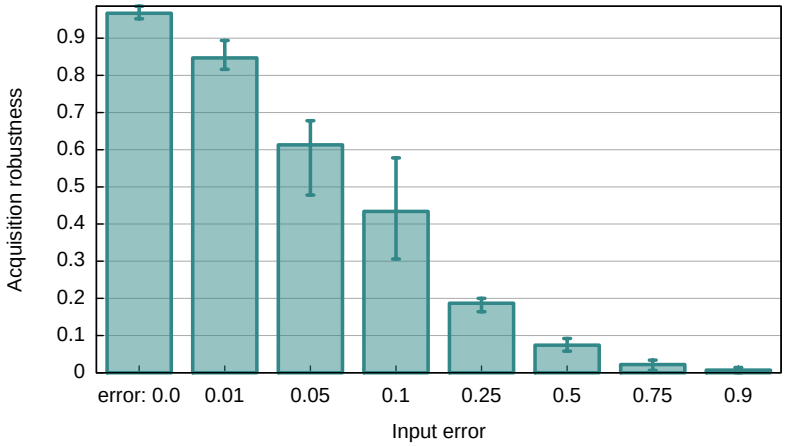


Figure 3.8: Acquisition robustness for the Baseline GG strategy with different $error_{input}$ values. Every bar represents a series of 500 exposures to the same word. Acquisition robustness represents the percentage of these in which the agent at the end associates the "correct" target meaning to the word. Number of simulations: 100, error: 5 to 95 percentile, context size: 10, total number of objects: 50.

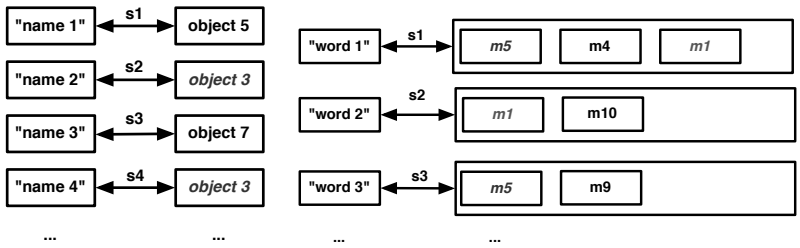
3.4 Cross-situational strategies

The Minimal Strategy in combination with a lateral inhibition scheme discussed in the previous Section can achieve full mapping alignment but does so at a very slow pace and only for small populations and small context sizes. The agents truly take guesses at the meaning of a word until they get it right. Since there is no agreed upon meaning to start from this is a painstakingly slow process. The representational capacities of Baseline GG agent is simply not up to the task of coping with uncertainty in systematic way.

3.4.1 The CS Set Strategy

The Cross-situational Set Strategy, or CS Set Strategy in short, extends the representational capabilities of the agents by allowing them to store multiple hypotheses per word as depicted in Figure 4.4(b).

In order to explain production and interpretation we require a function $f_{meanings}(w)$ which returns the associated set of meanings for word w and



(a) Naming Game memory (b) Cross-situational Guessing Game memory

Figure 3.9: (a) Bi-directional memory required for a Naming Game (or the Baseline GG Strategy). Mappings are scored and different names/words can refer to the same object. (b) Bi-directional memory required for the Cross-situational Set strategy for the Minimal Guessing Game. Multiple competing meanings can be associated with a word. The association between word and hypothesized meanings are scored in the same way as it was in more advanced Naming Game strategies.

$f_{meaning}(w)$ which returns a random element from $f_{meanings}(w)$. The production function $f_{produce}(o, a)$ looks up all words w for which $o = f_{meaning}(w)$. Note that if $o \in f_{meanings}(w)$ there is only a $\frac{1}{|f_{meanings}(w)|}$ chance that $o = f_{meaning}(w)$. When there is no such word $f_{invent}(o, a)$ creates a new word w_{new} and associates it with singleton $\{o\}$ since o is the only possible meaning for the speaker.

Given a word w and context C the listener a tries to interpret using $f_{interpret}(w, C, a)$ by looking up w. In case he does not find w in his lexicon $f_{adopt}(w, C, a)$ associates w with all the elements in C such that $f_{meanings}(w) = C$.

At the end of the game $f_{align}(w, C, a_{listener})$ updates the meanings associated with w as follows:

$$f_{meanings}(w) = \begin{cases} C & \text{if } f_{meanings}(w) \cap C = \emptyset, \\ f_{meanings}(w) \cap C & \text{otherwise.} \end{cases} \tag{3.1}$$

The agent takes the intersection of the possible meanings still associated with the word and the current context, as such reducing the associated meanings. When none of the associated meanings is compatible with the context, the agent starts over by associating all objects in the context. This update mechanism is related to simple candidate elimination algorithms known from concept learning (Mitchell, 1977). De Vylder (2007) also introduced a likewise

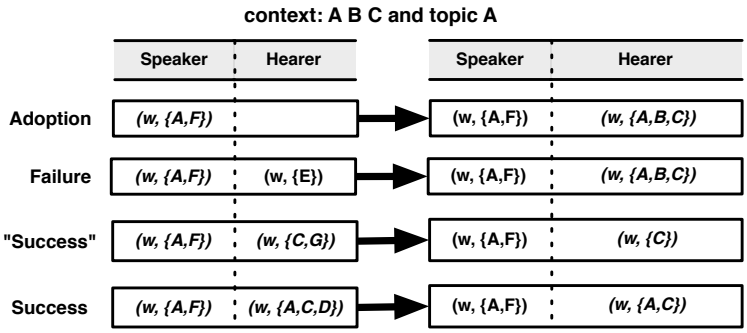


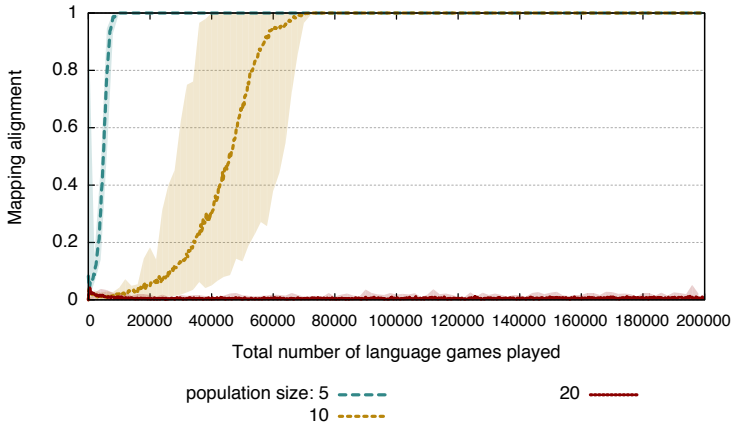
Figure 3.10: Update mechanics for the meaning of a word in the CS Set Strategy. Given context [A,B,C] and topic A (only known to speaker), the speaker utters the word w to which he still associates two meanings A and F. In adoption the listener does not know w. In case of failure the listener knows w but $f_{meanings}(w) \cap C = \emptyset$. In the third case, $topic \notin f_{meanings}(w)$ but $f_{meanings}(w) \cap C \neq \emptyset$. The listener reduces $f_{meanings}(w)$ but incorrectly since the intended meaning is not part of his associated meanings. Finally $topic \in f_{meanings}(w)$, in which case this set of meanings will be reduced to $f_{meanings}(w) \cap C$.

strategy, which he called the Naive Strategy and it has been investigated in depth by Smith *et al.* (2006) and Blythe *et al.* (2010). The update mechanics are schematically shown in Figure 3.10.

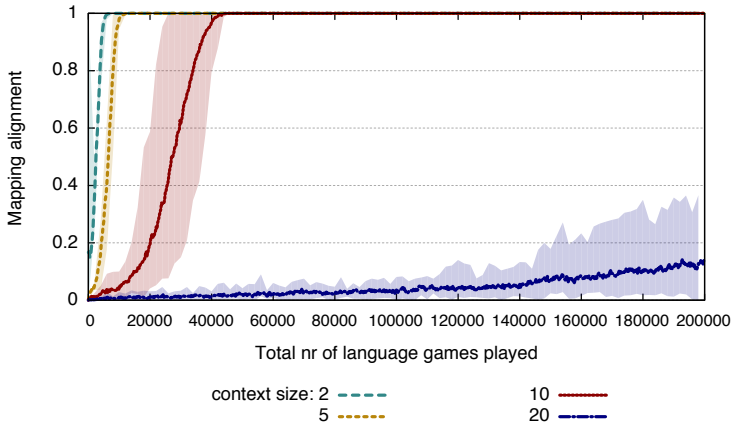
Only listeners and not speakers align because only in this role do they receive information about the preference of the other (speaker) agent. The goal of each agent is to arrive, for each word, at the same meaning as all other agents. From this perspective it makes little sense to change your lexical inventory based on your own production since it simply reflects your own preference and not in that of the other agent. This is the same reason as why in the Frequency NG Strategy from the previous chapter also only the listener updated frequencies.

We combine this strategy with the Interpolated Lateral Inhibition strategy from the previous chapter (Section 2.4, page 51) in order to dampen competing word forms. Like in the Interpolated LI Strategy for the Naming Game each word form has a word form competition score in the interval $[0, 1]$. This score is increased by δ_{inc} for the spoken word as follows:

$$\text{reinforcement:} \qquad s \leftarrow s + \delta_{inc}(1 - s) \tag{3.2}$$



(a) Scaling population size



(b) Scaling context size

Figure 3.11: Scaling of CS Set Strategy in combination with the interpolated Lateral Inhibition Strategy. Scaling is improved dramatically, especially with regard to context size. Larger population sizes still remain problematic. (a+b) total number of objects: 50, error: min/max, number of simulations: 12, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$. (a) context size: 4, population sizes: 5, 10 and 20. (b) population size: 5, context sizes: 2, 5, 10 and 20.

The score of all word form competitors is decreased by $delta_{inh}$:

$$\text{inhibition:} \qquad s \leftarrow s - \delta_{inh} s \tag{3.3}$$

A word w_c is considered a word form competitor of the spoken word w when $f_{meaning}(w_c) = f_{meaning}(w)$. The third parameter δ_{dec} is not used.

Figure 3.11 shows the mapping alignment of the strategy when scaling population and context size. Compared to the Baseline GG Strategy there is a drastic improvement in terms of alignment speed. Also with regard to scaling of context size the behaviour is acceptable. Scaling of the size of the population remains problematic. A population of 20 agents does not show the slightest bit of alignment after playing a total of 200000 language games.

As discussed in more detail by Blythe *et al.* (2010) the most important factor in determining the chances of success of convergence in a single game is the ratio of the size of the context C over the total number of objects O. They propose the fraction $\frac{|C|}{|O|}$ as a measure for referential uncertainty[2]. The larger $\frac{|C|}{|O|}$, the more difficult it is for a cross-situational strategy to learn. This makes sense because learning opportunities are greatest when overlap between two contexts is small. The smaller $|C|$ is compared to $|O|$ the smaller the chance for large overlap when picking two random contexts C_1 and C_2. In the extreme case where $|C| = |O|$ learning becomes impossible. See (Blythe *et al.*, 2010) for a more in depth investigation of the factors influencing uncertainty.

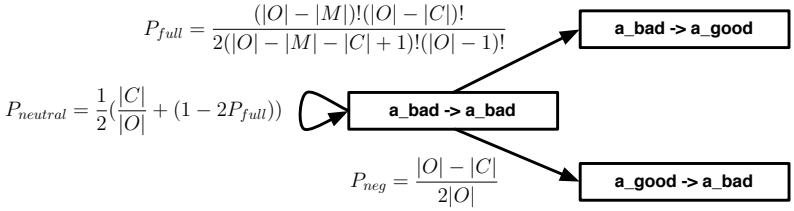
As we did for the previous strategy let us examine the chances of a single game in which an agent a_{good} interacts with a_{bad} and word w is spoken. Agent a_{good} already uniquely associates the conventional meaning o_1 to w, whereas a_{bad} still maintains a set of competing meanings M which does already contain the target meaning o_1[3]. Again we assume a context C and the set of all objects O. We distinguish three possible outcomes (the conditions and chances under which they occur are discussed immediately after):

(1) full alignment: The case in which a_{bad} prunes his set of competitors so that it only contains the target meaning $< w, \{o_1\} >$. *abad* thus becomes an a_{good}-type agent.

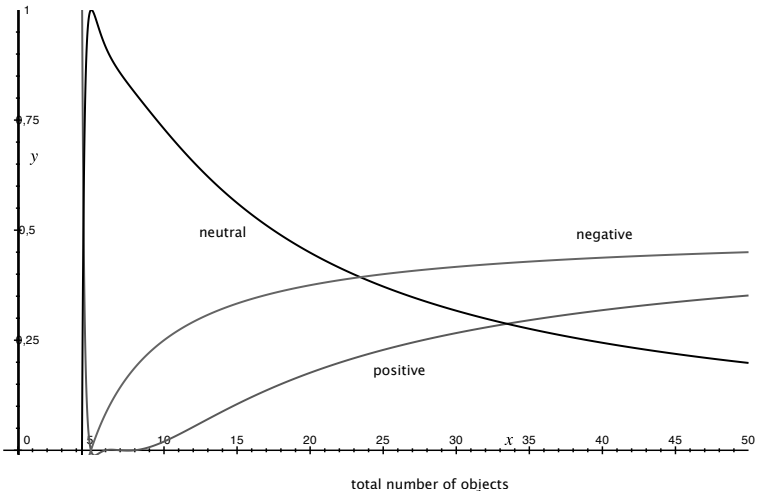
(2) neutral and positive: The case in which a_{bad} still maintains multiple competitors M, although their number might have shrunk. a_{bad} remains an a_{bad}-type agent. Given the a priori constraints it is impossible that the set of competitors would grow larger.

[2]Instead of O Blythe *et al.* (2010) use the symbol M to denote the set of all objects (or meanings in their case).

[3]If all other agents are like a_{good} then it would only take a single failed game with a_{bad} as listener to arrive at such a set of competitors, in case it did not yet contain o_1.



(a)



(b)

Figure 3.12: Alignment transition probabilities for the CS Set Strategy. (b) Parameters C and M are both taken to be 5. The total number of objects ranges from 5 to 50 (x-axis).

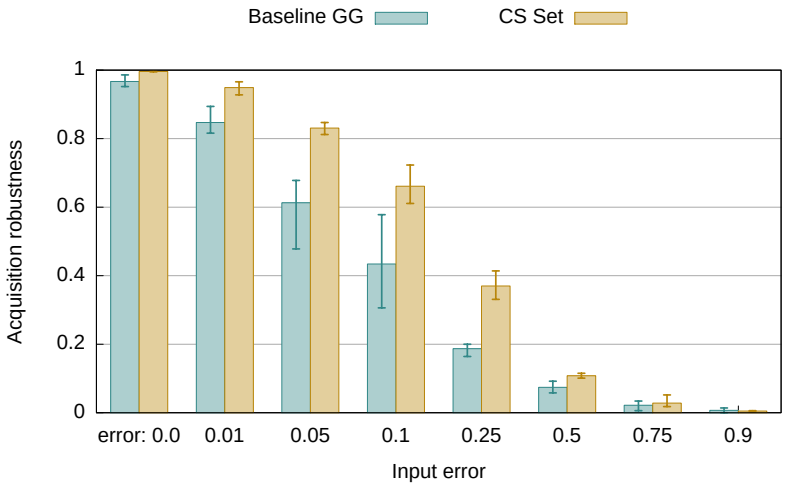


Figure 3.13: Acquisition robustness for the Baseline GG strategy and the CS Set Strategy with different $error_{input}$ values. Every bar represents a series of 500 exposures to the same word. Acquisition robustness represents the percentage of these in which the agent at the end associates the "correct" target meaning to the word. Number of simulations: 100, error: 5 to 95 percentile, context size: 10, total number of objects: 50.

(3) negative: The case in which a_{good} forgets the correct association and resets the set of competitors to C. a_{good} now becomes an a_{bad}-type agent.

The negative case is most easy to calculate as it requires a_{bad} to be speaker (chance $\frac{1}{2}$) and o_1 not to be part of the context C (chance $1 - \frac{|C|}{|O|}$) which amounts to a total chance $P_{neg} = \frac{|O|-|C|}{2|O|}$. The full alignment case requires a_{bad} to be listener and $M \cap C = \{o_1\}$ so that all elements in $M \setminus C$ are eroded. Since the word w is spoken by a_{good} we know that o_1 must be part of C so the number of contexts for which $M \cap C = \{o_1\}$ is $\binom{|O|-|M|}{|C|-1}$. The number of all possible contexts amounts to $\binom{|O|-1}{|C|-1}$ so that the chance to generate a context that leads to full alignment is $\frac{\binom{|O|-|M|}{|C|-1}}{\binom{|O|-1}{|C|-1}} = \frac{(|O|-|M|)!(|O|-|C|)!}{(|O|-|M|-|C|+1)!(|O|-1)!}$. This leads to $P_{full} = \frac{(|O|-|M|)!(|O|-|C|)!}{2(|O|-|M|-|C|+1)!(|O|-1)!}$. The neutral case is encountered when a_{bad} is speaker and o_1 is in the context which happens with a chance of $\frac{|C|}{2|O|}$ or when (which could also lead to a reduction in competitors) when a_{bad} is listener

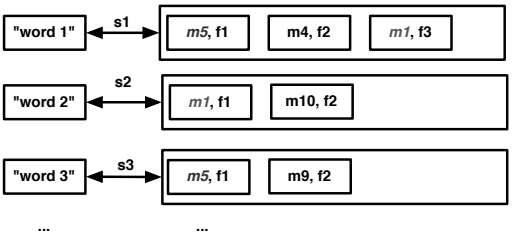


Figure 3.14: Bi-directional memory required for the Cross-situational Frequency Strategy for the Minimal Guessing Game. Multiple competing meanings can be associated with a word. Each associated meaning keeps the frequency of co-occurrence between the word and that particular meaning.

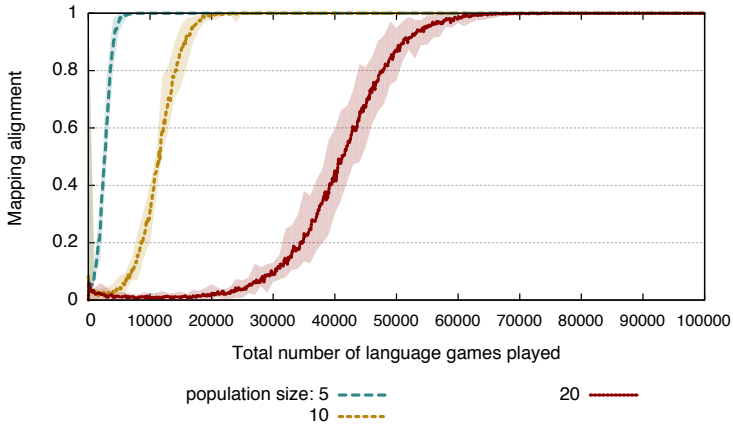
and $M \cap C \neq \{o_1\}$ which gives $P_{neutral} = \frac{1}{2}(\frac{|C|}{|O|} + (1 - 2P_{full}))$. See Figure 3.12 for a schematic depiction of the possible transitions and their chances.

The most important weakness of this strategy is that as soon as one non-compatible context is encountered the meaning is reset to the full context and cross-situational learning starts over. For acquisition experiments such as (Smith *et al.*, 2006; Blythe *et al.*, 2010) this is a reasonable prerequisite but in multi-agent emergence experiments inconsistent input is the rule, rather than the exception in the bootstrapping phase. This is shown in Figure 3.13 which shows that although the set-based CS Strategy improves on the non-cross-situational Baseline GG Strategy its acquisition robustness remains very poor.

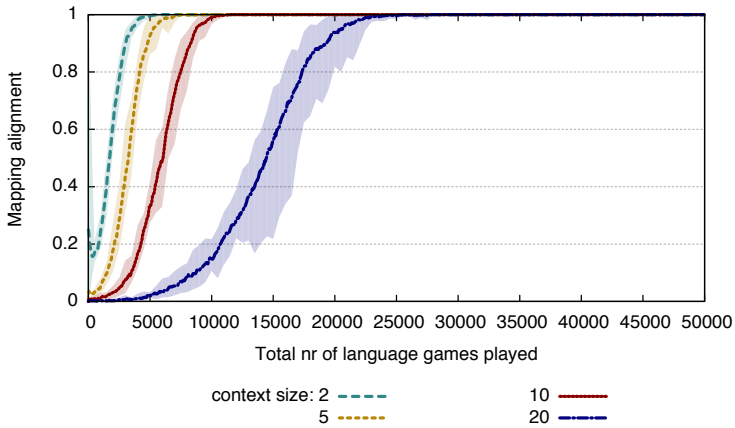
We now turn to strategies which increase their robustness by keeping frequencies or scores.

3.4.2 The CS Frequency Strategy

The Cross-situational Frequency Strategy, or CS Frequency Strategy in short, extends the representational capabilities of the agents further. Agents keep a frequency count for each meaning associated to a word. The frequency represents the number of times the meaning was part of, and thus compatible with, the context. The meaning with the highest frequency is then taken to be *the* meaning of the word (see Figure 3.14). This strategy exhibits behaviour very much like the Bayesian model proposed by Smith (2003a) and by Vogt & Coumans (2003) (in what he calls a selfish game). Here we combine the Frequency Strategy with a Naming Game Strategy in order to dampen word form competition.



(a) Scaling population size



(b) Scaling context size

Figure 3.15: Scaling of the Frequency Strategy in combination with the interpolated Lateral Inhibition Strategy. Scaling is again improved dramatically both in terms of population size and context size. (a+b) total number of objects: 50, error: min/max, number of simulations: 12, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$. (a) context size: 4, population sizes: 5, 10 and 20. (b) population size: 5, context sizes: 2, 5, 10 and 20.

The function $f_{meanings}(w)$ works the same as in the set-based strategy and returns all meanings associated with w. $f_{meaning}(w)$ returns the meaning with highest frequency, returning a random one if multiple meanings with the same frequency are associated. The function $f_{freq}(w, o)$ returns the frequency for the association of w and o. Production works uses the same algorithm as the set-based strategy by looking up all words w for which $o = f_{meaning}(w)$. Note however that we have redefined $f_{meaning}(w)$ to take into account frequencies. By using this new $f_{meaning}(w)$, $f_{produce}(w, o)$ takes these frequencies into account as well. When inventing a new word w_{new} the speaker associates his topic o with frequency 1 to w_{new}. In adoption the listener associates w_{new} with all objects in the context C, all with frequency 1. In alignment the listener updates the meaning of the spoken word $f_{meanings}(w) \leftarrow f_{meanings}(w) \cup C$ and the frequency scores follows:

$$f_{freq}(w, o) \leftarrow \begin{cases} f_{freq}(w, o) + 1 & \text{if } o \in f_{meanings}(w) \cap C \\ 1 & \text{if } o \in C \setminus f_{meanings}(w). \end{cases}$$

The reason only the listener updates is the same as explained for the Frequency NG Strategy and the CS Set Strategy.

Combining the CS Frequency Strategy with the interpolated Lateral Inhibition Strategy allows a population to reach mapping alignment as shown in Figure 3.15. As previously Figure 3.15(a) shows the performance when scaling the size of the population. Note that the number of games shown is no longer 200000 but only 100000 and 50000. In each of the settings the population reaches full mapping alignment. This is thus the first strategy that can handle all tested cases.

Since agents can keep track of frequencies they are much more resilient against inconsistent input. Figure 3.16 shows that acquisition robustness has increased dramatically. Even when half of the input is inconsistent a CS Frequency agent can maintain the correct association almost over all exposures. The reason is that the inconsistent input is itself not consistent and the inconsistent frequency counts are thus spread over all the objects. For example for a total of 50 objects, a context size of 10 and 500 exposures with $error_{input} = 0.5$ the frequency of the "correct" association would be on average 250 whereas for any other object it would be $250\frac{10}{50} + 250\frac{10}{49} = 101$.

The impact of a word form competition strategy

The word form competition problem, which is inherited from the Naming Game, might have been overlooked by some researchers in cross-situational learning. We have already looked at combining the Baseline GG Strategy

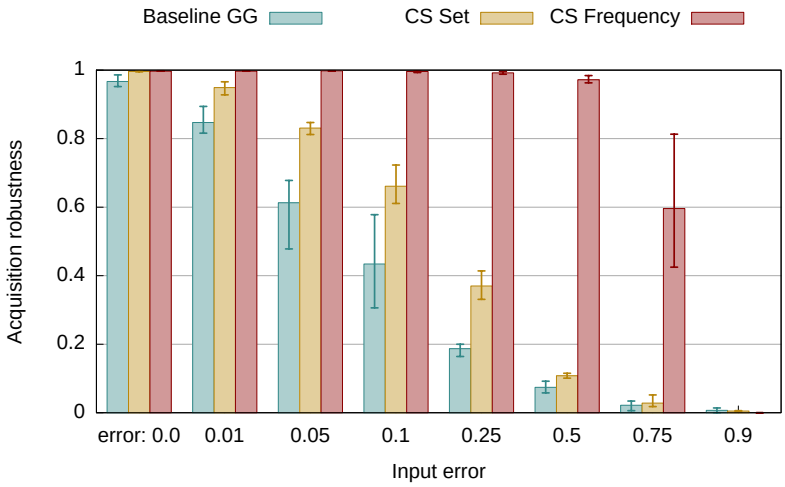


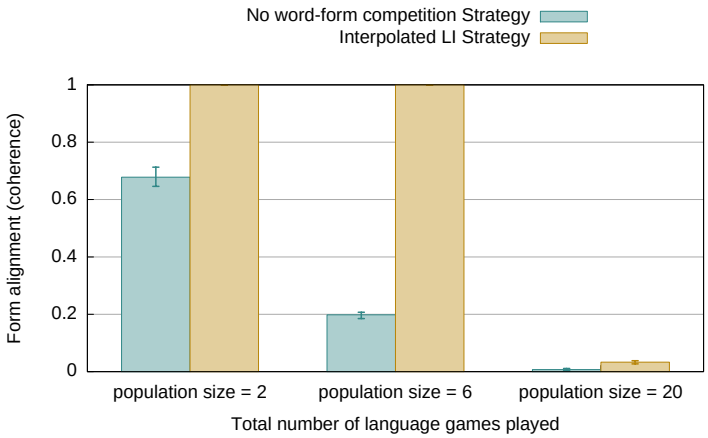
Figure 3.16: Acquisition robustness for the Baseline GG strategy, the CS Set Strategy and the CS Frequency Strategy with different $error_{input}$ values. Every bar represents a series of 500 exposures to the same word. Acquisition robustness represents the percentage of these in which the agent at the end associates the "correct" target meaning to the word. Number of simulations: 100, error: 5 to 95 percentile, context size: 10, total number of objects: 50.

with a Naming Game Strategy in Section 3.3, Figure 3.5. Here I would like to delve into this topic a bit deeper in combination with the Frequency Strategy.

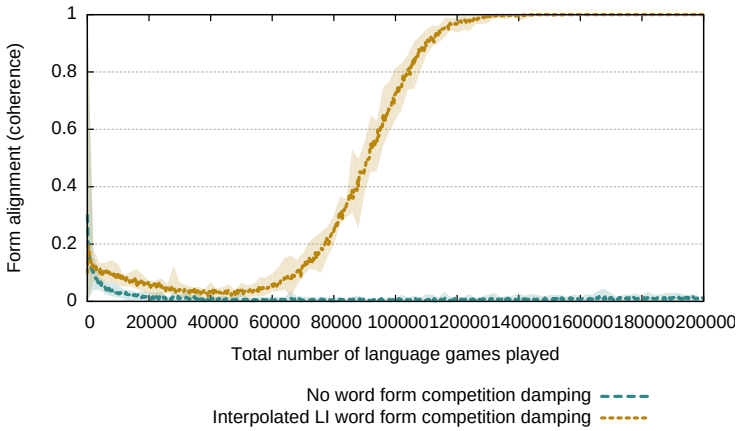
Smith (2003a) and Vogt & Coumans (2003) (in what they call the selfish game) introduce a Bayesian cross-situational strategy to cope with mapping uncertainty. In terms of dynamics this strategy is closest to the Frequency Strategy. The difference is that the score of a form-meaning pair $< w, o >$ in the Frequency Strategy is equal to the number of times w co-occurred with a context including o, whereas in the Bayesian Strategy this co-occurrence frequency is divided by the total number of times w was heard.

Vogt & Coumans (2003) also investigate the impact of increasing population size so I tried replicating his results[4]. Since I do not have a notion of communicative success I focus on the measure "coherence" Vogt & Coumans (2003) describe to measure whether agents produce the same word when given the same meaning. This measures what I call form alignment (see Section

[4]Since the strategies are not completely the same a 100% accurate one-to-one replication is impossible. I do believe that the dynamics are close enough so that the same overall dynamics and performance should emerge.



(a)



(b)

Figure 3.17: Impact of a NG Strategy on form alignment in combination with the Frequency CS Strategy. (a+b) total number of objects: 100, context size: 5, error: min/max, number of simulations: 10, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$. For the bar charts in (a) a total of 50000 games was played. (b) shows that if 200000 games are played word form competition does get eliminated for a population of 20 agents.

3.2).

Just like in (Vogt & Coumans, 2003) we take different population sizes (I only replicate for 2, 6 and 20), a total set of 100 objects (meanings) and 5 objects per context and let the population play 50000 language games over 10 different runs. In order to arrive at the bar charts reported by Vogt & Coumans (2003)[5] the average coherence is measured over the last 2500 games. In this experimental setup Vogt & Coumans (2003) report a coherence (form alignment) of 0.83 for 2 agents, 0.11 for 6 agents and 0.02 for 20 agents. This is a surprising result for two reasons: (1) why can even 2 agents not reach 100% coherence (form alignment) and why does it drop so drastically when the size of the population increases. This is not in line with the results shown in Figure 3.15a because it shows that (although for different parameters) a population can reach full mapping alignment (which is a stricter measure than form alignment).

I was, however, able to replicate the results but only when removing the damping algorithm for word form competition, only keeping the cross-situational learning and ignoring the Naming Game problems. Results with the same parameters as Vogt & Coumans (2003), using the same bar-chart measuring of the final 2500 games of 10 series of 50000 games are shown in Figure 3.17a. The bars showing the form alignment without an additional Naming Game Strategy map very well on those from Vogt & Coumans (2003) and can be seen as replicating the two findings that (1) two agents are not capable of reaching full form alignment (coherence) and (2) the measure drops dramatically with increasing population size. With an additional form-competition damping strategy (here Interpolated LI Strategy) a population of 2 and 6 agents can reach full form alignment in 50000 games. However for 20 agents and a total of 100 objects, 50000 games is just too little. In Figure 3.17b I do show that in the case more than 50000 games are played the population can reach full form alignment (coherence) as long as they also implement word form competition damping.

As a consequence I believe the cross-situational learning algorithm that Vogt & Coumans (2003) and Smith (2003a) propose is in fact more powerful than they conclude. The performance of their strategy is crippled[6] due to the lack of an added word form competition algorithm. Indeed, Vogt & Coumans (2003) observes that "It seems that each agent developed more or less its own language, which the other agents understood, although they did not converge in their production.". This observation is fully in line when a word form competition damping strategy is missing. This was shown for the Naming

[5] I am referring to Figure 5b in (Vogt & Coumans, 2003)

[6] It is fully crippled with respect to form alignment (coherence) and still slightly with respect to meaning alignment as was shown in Section 3.3.

Game in the previous chapter (see Section 2.3.3, Figure 2.4, page 2.4).

Weakness of the Frequency CS Strategy

One potential weakness of the Frequency CS Strategy is that the flexibility of the agents reduces over time as the frequencies keep on increasing. In other words the agents might become *too* robust and rigid. In order to switch from meaning m_1 to another meaning m_2 for a given form f_1 an agent needs to hear the pairing $< f_1, m_2 >$ at least one time more than the pairing $< f_1, m_1 >$. So if an agent heard the pairing $< f_1, m_1 >$ a thousand times and because of some external pressure the convention changes in the population to $< f_1, m_2 >$ it will take this agent at least 1001 exposures to start using the new convention.

De Beule *et al.* (2006) criticised exactly this aspect of the strategy and proposed an alternative where the agents maintain high levels of flexibility even after large numbers of exposures. This brings us to the Flexible CS Strategy.

3.4.3 The CS Flexible Strategy

De Beule *et al.* (2006) and De Vylder (2007) proposed a strategy which aims to improve upon the CS Frequency Strategy and make it more flexible at the same time. A strategy is flexible when its adaptivity does not reduce over time. The memory requirements of this strategy are similar to that of the Frequency Strategy except that the frequencies are replaced by scores in the interval $]0.0 - 1.0]$. $f_{meaning}(w)$ returns the meaning with the highest score associated to w, taking a random one if multiple meanings share the same highest score. Most processing, such as $f_{produce}$, $f_{interpret}$ and even f_{invent}, works in the same as in the Frequency Strategy. In invention the speaker only associates the topic as meaning with an initial score of 1.0. In adoption the listener associates all objects from the context C to the new word w_{new} with an initial score of $\frac{1}{|C|}$. Although the strategy presented by Smith (2003a) also maintains scores in the interval $]0.0 - 1.0]$ it has different update mechanics and shows a behaviour much more related to the Frequency CS Strategy.

The alignment function $f_{align}(w, C, a_{listener})$ which is only executed by the listener is more elaborate and quite intricate. Given the function $f_{score}(w, o)$ which returns the score of the association of w and o, the update rule then works as follows:

$$f_{score}(w, o) = \begin{cases} \beta(\gamma) f_{score}(w, o) \frac{\gamma'}{\gamma} + (1 - \beta(\gamma)) \frac{\gamma'}{|C|} & \text{if } o \in C \\ f_{score}(w, o) \frac{\delta'}{\delta} & \text{if } o \notin C. \end{cases}$$

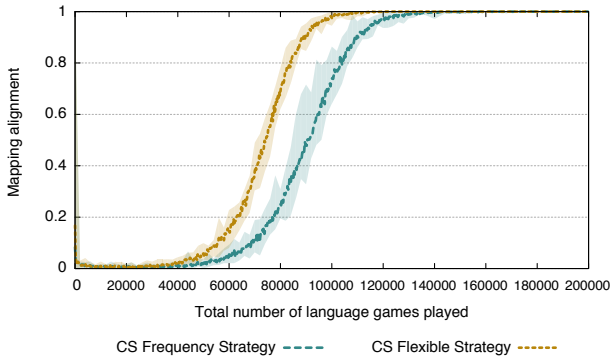


Figure 3.18: Comparison of mapping alignment for the Frequency and Flexible Strategy. Total number of objects: 100, context size: 5, total games: 200000, error: min/max, number of simulations: 12, population size: 20. $\delta_{inh} = \delta_{dec} = 0.3$, $\delta_{initial} = 1.0$ and $\alpha = 0.2$.

$\gamma = \sum_{o \in C} f_{score}(w, o)$ for the spoken word w which amounts to the total probability of all consistent objects in context C for word w. If *gamma* is close to 1 then the current context was highly compatible with the associated meanings, if it is close to zero it was incompatible. γ' is the new total probability we want to spread over the consistent meanings and is calculated as $\gamma' = (1-\alpha)\gamma + \alpha$ which means $\gamma' \geq \gamma$ depending on the learning parameter α. As you can see γ and γ' are only used when $o \in C$. β determines how γ' is spread over these new compatible objects. De Beule *et al.* (2006) propose that when γ is high (compatible context) the relative probabilities should stay more or less the same. This results in weak probabilities remaining weak and strong ones remaining strong. On the other hand if γ is low, which means this was a more disruptive (incompatible) exposure, γ' should be spread evenly. This is achieved by choosing $\beta(\gamma) = \sqrt{1-(1-\gamma)^2}$. This particular formulation also has the effect that as long as $gamma < 1$ all scores of objects in C increases. The learning parameter $\alpha \in [0.0 - 1.0]$ is taken to be 0.2 in all the reported experiments. Finally for the case $o \notin C$, $\delta = 1 - \gamma$ and $\delta' = 1 - \gamma'$. For an example I refer the reader to (De Beule *et al.*, 2006).

Just as with previous strategies we combine this strategy with the Interpolated Lateral Inhibition Strategy from the previous chapter with $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$. Figure 3.18 shows the mapping alignment of both the Frequency and the Flexible Strategy for a population of 20 agents, a total of 100 objects and a context size of 5 objects. Both strategies perform very alike with an edge for the Flexible Strategy in terms of speed of alignment.

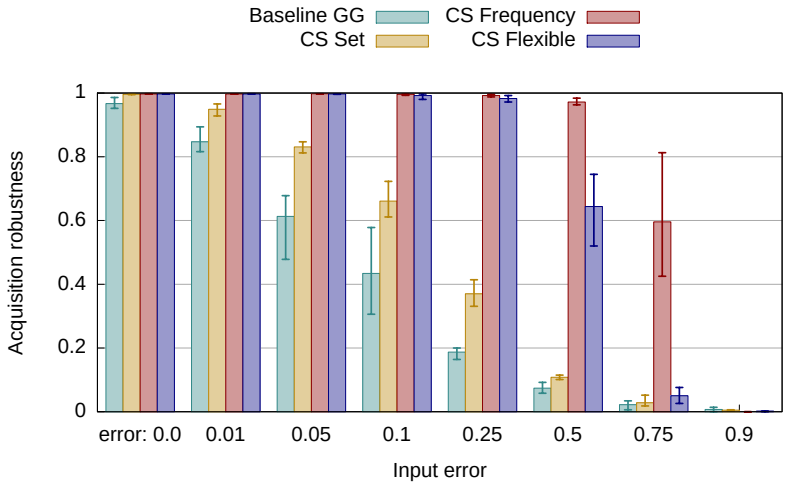


Figure 3.19: Acquisition robustness for the Baseline GG strategy, the CS Set Strategy, the CS Frequency Strategy and the CS Flexible Strategy with different $error_{input}$ values. Every bar represents a series of 500 exposures to the same word. Acquisition robustness represents the percentage of these in which the agent at the end associates the "correct" target meaning to the word. Number of simulations: 100, error: 5 to 95 percentile, context size: 10, total number of objects: 50.

This result also corroborates the results from De Beule *et al.* (2006).

Acquisition robustness for the CS Flexible Strategy is convincing in that even when 1 out of 4 exposures is inconsistent an agent can maintain the target association. It does break down at a lower $error_{input}$ than the CS Frequency Strategy does but this is exactly what the CS Flexible Strategy set out to do. The criticism of De Beule *et al.* (2006) was that the CS Frequency Strategy resulted in rigid agents that have trouble adapting to other emerging conventions. When the input is not correct half of the time it might indeed be a good idea to change your convention as maybe some of the other agents cannot or are not willing to change their preference. The CS Flexible Strategy thus tries to find a balance between robustness and flexibility.

	Naming Game	Minimal Guessing Game
meaning complexity	atomic	atomic
utterance complexity	single word	single word
word form competition	yes	yes
mapping uncertainty	no	yes

Table 3.1: Overview of Naming Game and Minimal Guessing Game. The Minimal Guessing Game differs only in one respect from the Naming Game in that it evokes mapping uncertainty.

3.5 Conclusion

This chapter discussed the problem of mapping words to atomic meanings or objects under uncertainty. A language user can no longer be certain about the intended meaning after only one exposure. Instead the learner needs to take into account multiple exposures of each word in different contexts or situations and (statistically) compare these contexts in order to infer the target meaning. This learning process is better known as cross-situational learning and in this chapter I have given an overview of three different cross-situational strategies.

I formulated this into the conventionalization problem of *mapping uncertainty* because agents need to establish a shared and minimal set of mappings between words and meanings. This is essentially the same lexical system as that of the Naming Game, only the road to reach this system is more difficult because of the added uncertainty. The language game that I propose to evoke this problem is a minimal extension to the Naming Game, which is why I call it the Minimal Guessing Game.

To show the need for cross-situational learning I also added a non-cross-situational strategy, the Baseline GG Strategy. In this strategy agents only keep a single mapping per word (one hypothesis) and keep this as long as there is no counter evidence for it (inconsistent context). Results showed that a population following this strategy can reach full mapping success but only when the population and the size of the context is kept small. The explanation lies in its utter fragility against inconsistent input.

I have also shown that this strategy does not solve the problem of word form competition, which is the conventionalization problem introduced by the Naming Game. In order to solve this problem the strategy needed to be supplemented with a Naming Game Strategy. It turned out that in combination with the mapping uncertainty problem, a scoring-based Naming Game strategy proved drastically better than the minimal variants. An observation which was not as clear when the word form competition problem was investigated in isolation (i.e. in Chapter 2).

	Baseline	Set-Based	Frequency	Flexible
(1) mapping alignment	yes	yes	yes	yes
(2) scaling	no	no	yes	yes
(3) robustness	no	no	yes	yes
(4) scoring	no	no	yes	yes
(5) solves WFC	no	no	no	no

Table 3.2: Overview of Minimal Guessing Game strategies. Following questions are addressed: (1) Can the strategy reach full mapping alignment for small populations and context sizes? (2) Does the strategy scale convincingly in terms of population and context size. (3) Is the strategy robust against inconsistent input. (4) Does the Strategy implement a scoring mechanism. (5) Does the strategy by itself also solve the problem of word form competition? If no, it means it requires the addition of a Naming Game strategy.

What the Baseline GG Strategy showed us is that there really is a need for a more powerful cross-situational strategy. I have reviewed three cross-situational strategies, which all have been proposed by other authors before.

The first strategy was introduced by Smith *et al.* (2006) and further investigated by De Vylder (2007) and Blythe *et al.* (2010). It implements cross-situational learning in what must be the most straightforward manner. It is no surprise De Vylder (2007) called it the "Naive" Strategy. Agents can keep for each word a set of competing meanings which they initialize to the complete context upon first exposure. Each new exposure reduces this set by removing non-compatible meanings. If all meanings would be removed (because of inconsistent input), the set is again initialized to the latest context. This strategy does not rely on scoring to keep track of cross-situational statistics.

Although the strategy shows an improvement upon the Baseline GG Strategy it still shows considerable issues when scaling population size or context size. The reason for this is that it still remains very susceptible to inconsistent input, which in a multi-agent language game setting happens frequently. This also explains why Smith *et al.* (2006) and Blythe *et al.* (2010) investigate this strategy from an acquisition point of view and under the assumption of consistent input. Indeed if input is consistent, Blythe *et al.* (2010) show that the Set-Based CS Strategy can scale to large lexicon sizes. The most important lesson here is that not any cross-situational strategy convincingly solves the mapping uncertainty problem.

I then show the results for two strategies that both allow agents to keep a score to track the cross-situational statistics in more detail. These strategies are based on work by Smith (2003a), Vogt & Coumans (2003), and De Beule

et al. (2006) among others. By implementing a cross-situational scoring mechanism both strategies show a much more robust behaviour against inconsistent contexts. These strategies, like all other strategies in this chapter, still do not solve the problem of word form competition. Just like the Baseline GG Strategy I again show that they need to be supplied with an extra word form competition damping strategy from Chapter 2. An overview of all strategies is given in Table 3.2.

The cross-situational strategies that I have covered here should not be taken as an exhaustive listing of all cross-situational models. For example more recently Fazly *et al.* (2010) introduced a probabilistic model for cross-situational learning which takes into account smoothing factors and also shows high levels of robustness to noise (inconsistent input). This is compatible with the findings here that cross-situational learners that track cross-situational statistics over longer periods (e.g. by using scores) can handle inconsistent contexts rather well.

Cross-situational learning has also been investigated in combination with other capabilities. For example Smith (2005b) found that adding the ability of mutual exclusivity to the algorithm not only reduces the initial set of competitors but also allows agents with different conceptual structures to communicate more successfully than those without this constraint. Mutual exclusivity means that when creating the initial set of competing meanings, only those for which the agent not yet knows a word are taken into consideration.

Chapter 4

Competitive strategies for Compositional Guessing Games

Chapter 2 introduced the Naming Game and with it the problem of name or word form competition. The Naming Game has two defining characteristics:

Atomic meanings: Meanings are assumed to have no internal structure, they are holistic and atomic.

Perfect feedback: Agents know the intended meaning of each word/name even on the first exposure.

We saw that even under these simplifying constraints a problem arises due to the multi agent nature of the language game paradigm. Different names or words are associated to the same meaning. This problem was called name or *word form competition*. Different strategies that tackle this problem were compared, amongst them the well-known strategies based on lateral inhibition.

Chapter 3 removed the assumption of perfect feedback and introduced the problem of mapping uncertainty. An agent could no longer be certain about the intended word-meaning mapping of a word upon first exposure. Meanings were assumed to be atomic and show no compositional structure. We saw that in order to solve the problem of mapping uncertainty in a robust and scalable way a cross-situational learning strategy capable of tracking certainties of competing meanings was required. Such a strategy had to be combined with a Naming Game strategy to dampen word form competition.

All case studies assumed (i) that word meanings are atomic and thus concerned with at most a single semantic domain and (ii) that an utterance contains only a single word. Obviously none of these restrictions holds for human natural languages. Many words express bundles of categories and thus incorporate many different semantic domains. Moreover almost all utterances are compositional: More than one word is used to express the

set of categories that the speaker intends to convey. Human languages are therefore compositional (different words express meaning sets that can be combined into a compositional utterance) rather than holistic (a single word expresses all of the meaning).

Compositionality and multi-dimensionality are intertwined because if one word can express many categories and multiple words can be used, speakers need to decide how to divide up concepts over different words, and listeners are faced with the problem of finding out the part of meaning expressed by each word. For example, suppose the speaker wants to express [tall green bright block]. He may choose to do this holistically with a single word, say "dobido", that expresses all these concepts, or with two words, for example, "bado" meaning [tall green] and "zobo" [bright block]. But how can the listener then know that "bado" does not mean [tall green bright] and "zobo" [block] or any other combination of possibilities? The problem grows exponentially, as opposed to linearly. For example, an object represented by twenty binary features leads to over one million possible subsets. Clearly a naive solution will not do.

In this chapter we look at the Compositional Guessing Game which introduces lexical compositionality and multi-dimensionality in the language game paradigm. In the Compositional Guessing Game objects are represented as sets of *attributes*. Besides the set of all objects $O = o_1 \cdots o_n$ from which contexts $C \subset O$ are drawn the Compositional Guessing Game adds the set of all attributes $A = a_1 \cdots a_m$ from which the objects $o_n \subset A$ are composed. Meanings M no longer refer to objects as a whole or single atomic attributes but can encompass any set of attributes $M \subset A$.

In the previous chapter lexicon learning and emergence was a matter of establishing a shared mapping from a set of forms F and a set of meanings or objects M as depicted in Figure 4.1. With the addition of attributes the metaphor of mapping form to meaning becomes less appropriate since the meanings themselves, which are now sets of attributes are not given from the start. The agents thus need to establish a shared set of meanings (attribute sets) and the forms they will use to express those. The uncertainty is no longer one of making the correct mapping but about finding out what is constitutive for the meanings of the emergent lexicon, which is why it is no longer called mapping uncertainty but *meaning uncertainty*. This scenario is depicted in Figure 4.2.

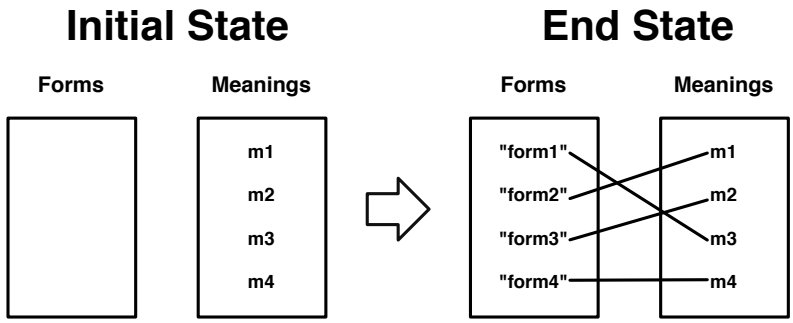


Figure 4.1: In the Minimal Guessing Game the agents need to establish a shared mapping of forms to meanings. The meanings are assumed to be atomic and are already given.

4.1 Differences with debate on the emergence of compositional or holistic languages

The Compositional Guessing Game is thus compositional in more than one sense. First the objects and the meanings are compositional or multi-dimensional. Second also the utterances can consist of multiple words, making them compositional as well. There has been quite a large body of research regarding the origins of compositionality in language, including from a computational modeling perspective (Kirby, 2000; Nowak M.A., 2000; Brighton, 2002; Vogt, 2005b; De Beule & K. Bergen, 2006). The main question that is addressed is how people started to combine multiple words to express and relate in a meaningful way multiple concepts. For example the phrase "Luc eats an apple" introduces an eating event, the concepts of an apple and a referent Luc through the lexical items "eats", "apple" and "Luc". Furthermore the addition of grammatical form constraints such as word order and subject-verb agreement further inform an English speaker that Luc is the actor of the eating event, and that an apple is what is being eaten. In a more logical notation this meaning could be written as [(Luc l) (eating e) (actor e l) (apple a) (undergoer e a)], although other notations are possible as well. Since they are interested in the emergence or evolution of compositional structure they are also interested in its counterpart, namely a more holistic language system in which lexical items express chunks of concepts. Regarding the example above the most extreme case would be a single word to express that entire meaning.

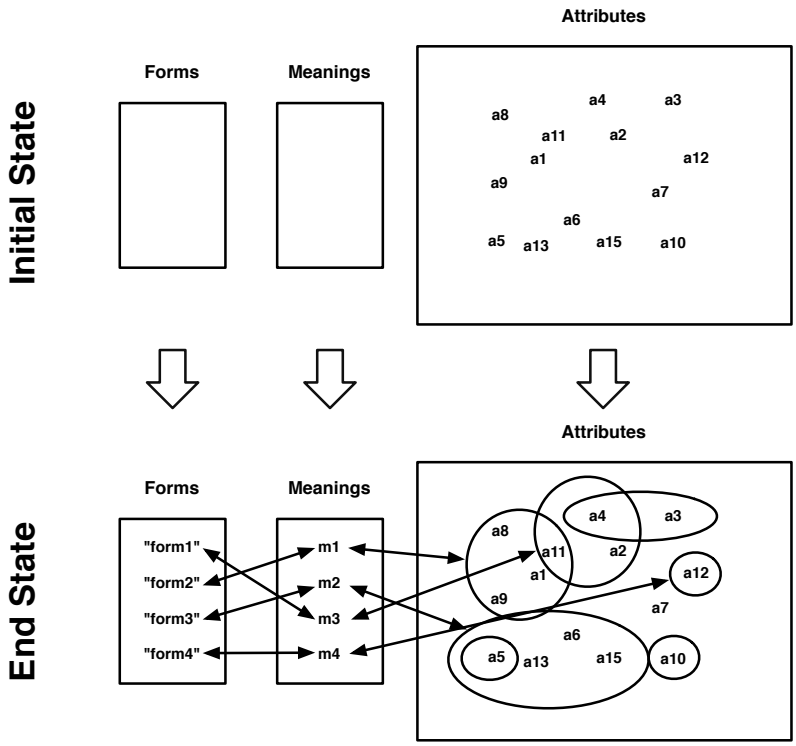


Figure 4.2: In the Compositional Guessing Game agents need to establish a shared set of meanings and at the same time need to map these meanings to words. This is the problem of meaning uncertainty.

Although there is an overlap between this work on compositionality and the work I present in this and the next few chapters, it would be a mistake to see the two as addressing exactly the same issue or even the same conventionalization problem. For one, in this and the following two chapters there is no syntax or grammar and even in Chapter 7 the "grammar" is much more embryonic in that no relational meaning is expressed (e.g. grammatical relations such as actors or undergoers). But most importantly, in the works cited above the primitive elements of the meaning representation are assumed to already be concepts and thus reside at a higher level of granularity than the primitive elements in the Compositional Guessing Game. In the Compositional Guessing Game the primitive elements are attributes out of which word

meanings (or concepts) need to be formed. As such the conventionalization problem is a different, yet partially related, one.

The fundamental difference is that the problem of meaning uncertainty is a lexical conventionalization problem about how meanings of individual words are formed. If concepts are however taken to be the primitives of representation then this problem is already assumed to be solved. So the core problem addressed in (Kirby, 2000; Nowak M.A., 2000; Brighton, 2002; Vogt, 2005b; De Beule & K. Bergen, 2006) either assumes that the problem of meaning uncertainty has already been solved or that it never existed in the first place.

Overlap is found in the use of multiple word utterances and the potential for internally structured meanings. In the work on compositionality a non-atomic meaning is, however, seen as not having reached full compositionality and the goal is to investigate strategies that arrive at a fully or at least highly compositional language, consisting mainly of words expressing atomic meanings (because their meanings are concepts), combined with a sort of grammar. The problem addressed in this and the following two chapters, namely the problem of meaning uncertainty depicted in Figure 4.2, is very different from that. Here the goal is not, or at least not a priori, to arrive at a fully atomic lexical system but instead arrive at a lexical systems that shows a richness in the meanings it covers, ranging from atomic (general) to multi-attribute (more specific), preferably in a usage-based manner. This will become clearer when discussing and evaluating the strategies.

4.2 Script of the Compositional Guessing Game

The compositional meaning specification allows for a more intuitive language game script (see Figure 4.3).The Compositional Guessing Game is played by two agents, one speaker, one listener. Both perceive a context C consisting of n objects o_i. Each object is composed out of k attributes a_j. The speaker picks one of these objects as the *topic* and constructs an utterance, possibly consisting of multiple words, which he believes will draw the attention of the listener to that object. Depending on the strategy the speaker can, but is not required to, take the other objects in the context into account. One way to take the context into account is to express the most discriminating attributes of the topic with regard to the other objects in the context.

The listener parses the utterance to its meaning and points to the topic which he believes corresponds most to the interpreted meaning. When pointed to the correct object the speaker acknowledges this by signalling success, if not the speaker points to the intended topic. Note that the speaker only points to the object, never to the attributes that were actually expressed.

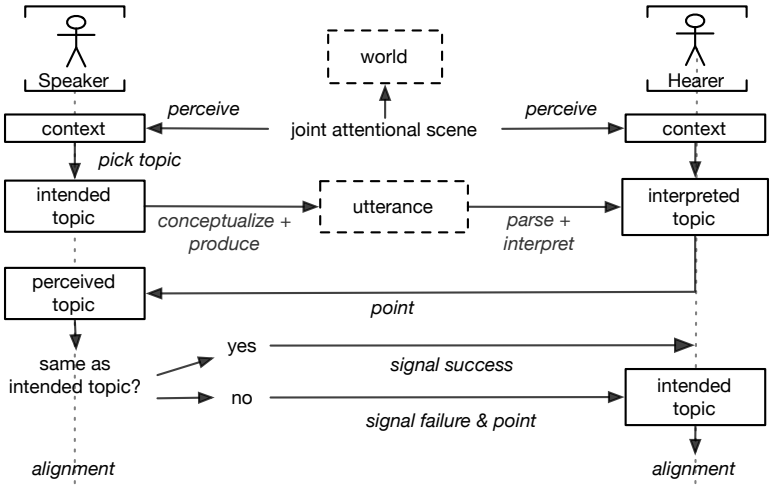


Figure 4.3: Script of the Compositional Guessing Game. Squared boxes represent data entities. Italicised text refers to processes. The production and interpretation processes can fail and therefore have diagnostics and repair strategies associated with them. A language game experiment consists of thousands of these interactions.

The meanings of the words can be any subset of the topic-attributes. The game is considered a success when the listener points to the correct object.

In this chapter we attempt to re-use and extend the cross-situational strategies from the previous chapter and proposed in Siskind (1996); Smith (2005a); De Beule *et al.* (2006); Smith *et al.* (2006); Vogt & Divina (2007); Van Looveren (1999) to the problems introduced by multi-dimensionality and lexical compositionality. These models were however not developed to handle a communicative task as demanding as the Compositional Guessing Game. For example most of them (Siskind (1996); Smith (2005a); Smith *et al.* (2006); Vogt & Divina (2007)) assume atomic meanings as in the Minimal Guessing Game and thus do not allow combinations of attributes. Only Van Looveren (1999), Steels *et al.* (2002) and De Beule (2007) extended the model to cope with set-based meanings.

4.3 Representational aspects of competitive strategies

The cross-situational strategies from the previous chapter have in common that they represent uncertainty by enumerating competing hypotheses. Each word lists all its potential meanings, and based on new exposures candidates are eliminated until only one meaning remains. Candidate meanings thus *compete* with each other until only one meaning remains. For this reason I will call these strategies *competitive*.

Two competitive strategies are introduced in this chapter. Both strategies share the same representational capabilities and thus differ only in processing. At the representational level they keep true to the core aspects of a competitive cross-situational strategy by allowing agent to enumerate competing hypotheses. More specifically they can also track frequencies for each word-meaning pair $< w, m >$ in exactly the same manner as the Frequency CS Strategy discussed in the previous chapter (Section 3.4.2, page 88).

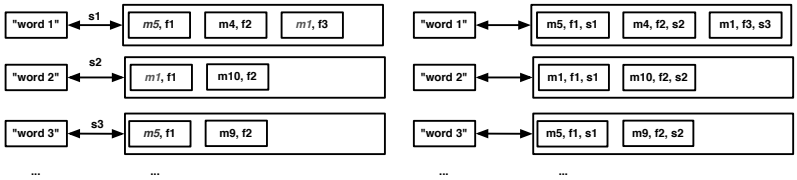
In short this means that for each word form an agent can enumerate all its associated competing meanings with the function $f_{meanings}(w)$. What of course has changed compared to the Minimal Guessing Game is that the meanings are sets of attributes instead of individual primitives. Function $f_{freq}(w, m)$ counts the number of times meaning m was compatible with the topic whenever w was heard. Since meanings and objects are both sets compatibility amounts to whether m was a subset of the topic t. For example meaning $\{a, b\}$ is compatible with topic $\{a, b, c, d, e, f, g\}$ but meaning $\{a, b, z\}$ is not. The frequency is thus, just like in the Frequency CS Strategy, keeping cross-situational statistics among competing meanings. Function $f_{meaning}(w)$ returns the meaning $m \in f_{meanings}(w)$ that maximizes $f_{freq}(w, m)$. If multiple such meanings exist one is chosen at random.

Each word-meaning association also maintains a word form competition score $f_{score}(w, m)$ which is updated by the interpolating lateral inhibition scheme as explained in Chapter 2, Section 2.4, page 51. In the previous chapter I chose to replicate the earlier models (e.g. De Beule *et al.* (2006)) as close as possible in this regard and thus chose the following definition of a word form competitor:

- A word form w_c is taken to be a word form competitor of word form w when $f_{meaning}(w_c) = f_{meaning}(w)$.

There is a potential flaw in this definition and I will illustrate it through the following example.

1. Assume at time t word w_1 and w_2 are word form competitors because $f_{meaning}(w_1) = f_{meaning}(w_2)$, and assume this meaning is m_1.



(a) Naming Game memory

(b) Cross-situational Guessing Game memory

Figure 4.4: (a) Bi-directional memory required for the Frequency CS Strategy for the Minimal Guessing Game. Frequencies f_x are kept per pair $< w, m >$ but word form competition scores s_n are kept at a higher level. (b) Modification of the Frequency CS memory to also keep the word form competition score at the level of the individual word meaning pairs $< w, m >$.

2. Assume that after playing many more games at a later time t', $f_{meaning}(w_2)$ is no longer m_1 but m_2 because $f_{freq}(w_2, m_2) > f_{freq}(w_2, m_1)$.

3. From this change it follows that w_1 and w_2 are no longer word from competitors because $f_{meaning}(w_1) \neq f_{meaning}(w_2)$. In fact w_2 is now a word form competitor of all words w_i for which $f_{meaning}(w_i) = m_2$.

4. The word from competition score of w_2 now makes no sense any longer since it is now competing with different word forms. Keeping the same definition it should at least need to be re-initialized to its initial value $\delta_{initial}$.

5. This is not an exceptional series of events but in fact what is happening frequently during bootstrapping of the lexicon.

To alleviate this flaw I have fine-tuned the word form competition damping by not keeping $f_{score}(w, m)$ per word form w but for each individual word meaning pair $< w, m >$. The word from competition score has thus moved to the same level as the frequency score $f_{freq}(w, m)$ as shown in Figure **??**b. The Interpolated Lateral Inhibition Strategy thus dampens for each produced or interpreted word meaning pair w, m all competing word meaning pairs w_c, m, which are all pairs that differ in form but share the same meaning m. As such word form competition is decoupled from function $f_{meaning}(w)$ and the problem sketched above is avoided. This will be further explained when discussing alignment.

4.4 Processing components of the Baseline Competitive Strategy

4.4.1 Production

The task of the speaker is to draw the attention of the listener to the object he chose as the topic. In order to reach this goal the speaker can utter one or several words each expressing one or more attributes. Word meanings are also allowed to partly overlap. In the baseline strategy the speaker does not take the other objects in the context into account from the very start and just tries to express as many attributes of the topic as possible.

The production function relies on a reduced version of his lexicon called the *usage lexicon*. The idea is that the usage lexicon shows only the form meaning pairs that, based on their scores, can be used during production. To reach that goal it has the two following modifications compared to the complete lexicon:

- For each word w in the lexicon only $f_{meaning}(w)$ is taken as its meaning (no meaning competitors).

- When for two words w_i and w_j it holds that $f_{meaning}(w_i) = f_{meaning}(w_j)$ then only the word which has the highest form competition score $f_{score}(w, f_{meaning}(w))$ is kept (no form competitors).

Calculating the usage lexicon does not modify the actual (complete) lexicon, it only takes the subset a speaker agent is interested in during production. If an agent has multiple words to express the same meaning he should only consider his current preference and the same reasoning applies to multiple competing meanings associated to a single word form. The main motivation to introduce the notion of a usage lexicon is to keep the production algorithm clean.

The production function constructs an utterance by expressing as many attributes of the topic as possible, giving preference to words that express multiple attributes at once. For example when the topic consists of attributes [red tall rectangular 3d shape] and the speaker has three compatible words ⟨"rood",{red}⟩, ⟨"rechthoek",{rectangular}⟩ and ⟨"blok",{rectangular, 3d, shape}⟩ the production function constructs utterance "blok rood". "blok" comes first because it expresses most attributes (three in this case), "rechthoek" brings nothing new and "rood" expresses one other attribute. Since production uses the usage lexicon both $f_{freq}(w, m)$ and $f_{score}(w, m)$ are taken into account. Pseudocode is shown in Algorithm 1.

```
———————— Function ProduceCompetitive(topic, agent) ————————
lexicon ← UsageLexicon(agent) sorted on Length(f_meaning(w));
// Words that express more attributes are at the front of lexicon
utterance ← empty;
Loop
  expressedMeaning ← ⋃_{w ∈ utterance} f_meaning(w);
  newCandidateWord ← empty;
  ForEach <word,meaning> pair in lexicon do
    newMeaning ← expressedMeaning ∪ f_meaning(word);
    If |newMeaning| > |expressedMeaning|
       AND newMeaning ⊆ Attributes(topic)
    then newCandidateWord ← word;
      and exit ForEach;
  End ForEach
  utterance ← utterance ∪ newCandidateWord;
while there is a newCandidateWord;
Return utterance;
```

Algorithm 1: Function ProduceCompetitive(topic, agent) expresses the topic in the context.

4.4.2 Interpretation

The listener has to point to one object in the context to which he believes the utterance is referring. In the Competitive Strategy the agent parses the utterance to its meaning by taking the union of the preferred meanings $f_{meaning}(w_i)$ of the words w_i in the utterance and points to the object for which the parsed meaning is a subset of its attributes. When there is not exactly one such object in the context interpretation fails. Pseudocode for the interpretation algorithm is given in Algorithm 2.

Both production and interpretation can fail, potentially triggering the agent to extend its lexicon. This is where invention and adoption come in.

4.4.3 Invention and adoption

The flow of the language game depicted in Figure 4.3 does not include invention and adoption of new lexical forms during production and parsing respectively. In addition, even when production and parsing were successful the game itself might be a failure (e.g. because of misaligned lexicons), again forcing the agents not only to adapt their scores but also to alter their linguistic inventories more drastically.

At the end of production, before actually uttering the utterance, the speaker first places himself in the role of the listener and tries to interpret

```
———— Function Interpret(utterance, context, agent) ————
interpretedMeaning ← ⋃_{w ∈ utterance} f_meaning(w);
ForEach object in context do
  If interpretedMeaning ⊆ Attributes(object)
  then push object onto possibleReferents;
End ForEach
If thereis only one possible referent;
then point to that referent;
else signal failure;
```

Algorithm 2: Function Interpret(utterance, context, agent) compares the best parse with each object in the context. This best parse is made from the union of the preferred meanings of each word in the utterance.

the utterance himself, a process called *re-entrance* (Steels, 2003) (also see Section 1.3.2, page 10). When re-entrance leads the speaker to a different object than his own intended topic it means that, at least for the speaker, no utterance could be found which successfully describes the topic in the current context. Without a new word the speaker cannot talk about the topic in the given context and the situation thus calls for a refinement of his lexicon. The speaker invents a new word form (a random string) and associates to it all attributes of the topic not yet expressed by other words in the utterance (see Algorithm 3). This invention strategy guarantees that the speaker can now express the topic in this context. After invention the speaker tries to produce again.

```
——— Function Invent(utterance, topic, context, agent) ———
// Diagnosing the utterance through re-entrance
interpretedObject ← Interpret(utterance, context, agent);
If interpretedObject = topic
then Return utterance;
else // Extend lexicon
  expressedMeaning ← ⋃_{w ∈ utterance} f_meaning(w);
  newMeaning ← Attributes(topic) \ expressedMeaning;
  addNewWord(randomForm,newMeaning,agent);
  ProduceCompetitive(topic, agent); // Try again
End If
```

Algorithm 3: Function Invent(utterance, topic, context, agent) diagnoses whether a new word should be added. If so a new word is created with the unexpressed attributes of the topic as its meaning.

The listener might encounter one or more novel words in the utterance.

In this case the listener ignores the novel words, interprets the words he does know and tries to play the game without adopting the novel forms. Only at the end of the game, when he has knowledge of the topic (see Figure 4.3), does the listener try to adopt the new word forms. In case there is more than one unknown word the listener will only adopt the first novel word of the utterance. The main reason for only adopting a single word at a time is because otherwise all unknown words would start out with the exact same set of competing hypotheses. If the agent only adopts one word at a time two words adopted with the exact same set of hypotheses should be rare. Furthermore it also allows an agent to learn words slightly more incremental. That the first word is picked does not have any significance, it could just as well have been any other unknown word.

The listener cannot know the intended meaning of this new word as it can encompass any subset of attributes of the topic. In order to reduce the initial set of hypotheses slightly the listener takes into account the parsed meaning of the words he does know and assumes the intended meaning of the word should be found among the remaining unexpressed attributes. For example, assume the listener needs to interpret utterance $w_1 w_2$ with topic $[a, b, c, d]$ and only knows word w_1 with $f_{meaning}(w_1) = [a, c]$. In this case he will assume that w_2 can only refer to the attributes b and d, resulting in only three initial competing meanings $[b]$, $[d]$, and $[b, d]$ instead of $2^4 - 1 = 15$. In this strategy we thus assume the constraint of mutual exclusion of meaning (Markman *et al.*, 2003), but only in learning (adoption and invention), not in processing (meanings can overlap). In the baseline Competitive Strategy the listener acts more or less the same as in the cross-situational strategies of the previous chapter and associates every possible remaining meaning with the novel word form. Pseudocode can be found in Algorithm 4.

```
Function Adopt(utterance, topic, agent)
If unknown forms in utterance; // diagnose
then // repair
  novelForm ← first novel form in utterance;
  interpretedAttributes ← ⋃_{w_i ∈ utterance} f_meaning(w_i);
  unexpressedAttributes ← topic \ interpretedAttributes;
  ForEach non empty subset newMeaning in unexpressedAttributes do
    AddWord(novelForm,newMeaning,agent);
  End ForEach
End If
```

Algorithm 4: Function Adopt(utterance, topic, agent) diagnoses for novel word forms and if so adds hypothesized meanings based on the words already known by the agent.

4.4.4 Alignment

With invention and adoption covered this leaves us with the alignment strategy. Alignment is the most crucial part of the language game because it implements the necessary self-organizing dynamics. As in the previous chapter only the listener aligns at the end of a guessing game and he will only do so in case he did not already adopt a new word.

The goal of alignment is to slightly alter the lexicon so that it is better aligned with that of the speaker. At the moment of alignment the listener has knowledge of the intended topic (through pointing), the utterance and the context. In the baseline Competitive Strategy alignment consists of three steps, (1) damping of word form competition, (2) updating frequency scores and (3) adding new hypothesized meanings.

The problem of word form competition remains the same as it was in the Naming Game where this problem was first encountered. As explained in Section 4.3, function $f_{score}(w, m)$ keeps a score in the range $]0, 1[$ for each word meaning association. This score is used in calculating the usage lexicon and thus influences all linguistic processing. Two words are seen as competitors when their $f_{meaning}(w)$ is equal. For each word-meaning pair the listener used, $f_{score}(w, m)$ is reinforced and that of all word form competitors is inhibited with the interpolating Lateral Inhibition Strategy (see page 51, Section 2.4, Chapter 2) as follows:

$$\text{reinforcement: } f_{score}(w, m) \leftarrow f_{score}(w, m) + \delta_{inc}(1 - f_{score}(w, m))$$
$$\text{inhibition: } f_{score}(w, m) \leftarrow f_{score}(w, m) - \delta_{inh} f_{score}(w, m)$$

In all experiments $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$. This is the same solution that proved successful for the cross-situational strategies discussed in the previous chapter. Code is shown in step (1) in Algorithm 5.

Frequency scores are updated for every compatible word meaning association in the current game. More specifically, for each spoken word w, each associated (competing) meaning in $f_{meanings}(w)$ is checked whether its set of attributes is a subset of the attributes of the topic. If so it is compatible and $f_{freq}(w, m)$ is incremented by one. This corresponds to step (2) in Algorithm 5.

In the case that communication is unsuccessful (the listener pointed to the wrong object), and all words are known to the listener, the listener extends the list of competing meanings for each word. For each spoken word w the listener checks whether any of the associated meanings $f_{meanings}(w)$ is a subset of the topic. Only when no meaning proves compatible will the listener extend the list of meanings for word w. For this he uses the same mechanism as in adoption. The listener thus takes all known words into

```
———————— Function Align(usedWords, topic, agent) ————————
ForEach word-meaning pair <w,m> in usedWords do
  // (1) Dampen word form competition
  f_score(w,m) ← f_score(w,m) + δ_inc(1 − f_score(w,m));
  ForEach word-meaning pair <cw,cm> in FormCompetitors(w,m) do
      f_score(cw,cm) ← f_score(cw,cm) − δ_inh f_score(cw,cm);
  End ForEach

  // (2) Update frequencies
  ForEach meaning in f_meanings(word) do
    If meaning ⊆ Attributes(topic)
    then Increment(f_freq(word,meaning));
  End ForEach
End ForEach

If Communication was unsuccessful
then // (3) Update meaning hypotheses
  ForEach <word, meaning> pair in usedWords do
  If ∀ m ∈ f_meanings(word): m ⊈ Attributes(topic)
  then // Add new hypotheses
    expressedAttributes ← ⋃_{word ∈ (usedWords \ word)} f_meaning(word);
    unexpressedAttributes ← topic \ expressedAttributes;
    AddWord(word,unexpressedAttributes);
  End ForEach
```

Algorithm 5: Function Align(usedWords, topic, agent). See text for details.

account and associates every subset of the unexpressed attributes as a new candidate meaning. Pseudocode can be found in step (3) of Algorithm 5.

4.5 Experimental setup and measures for the Compositional Guessing Game

Experiments in the Compositional Guessing Game have a few important parameters. As before the size of the population remains important since it strongly influences the rate of invention and the speed of convergence. The size of the context, which in the Minimal Guessing Game is the main regulator of uncertainty is of lesser importance in the Compositional Guessing Game since there is corrective feedback (pointing) at the end of the game. Uncertainty is now regulated primarily by the number of attributes per object since it determines the number of potential meanings upon hearing a novel form. The total number of attributes and the statistical distribution of the attributes over the objects also plays a role, which will be investigated later

on.

In the experiments we can regulate these parameters and as such investigate the performance of a strategy under different conditions and difficulties. First and foremost, we are interested in the *communicative success* the agents reach in playing the Compositional Guessing Game. As before communicative success is a Boolean measure which is true (1) only when the listener pointed to the intended topic.

Besides communicative success, we are also interested in the level of alignment of use of the emergent lexicons, a measure called *lexical alignment*. This measure compares between both participating agents at the end of each game for each word in the utterance the similarity of the associated meanings $f_{meaning}(w)$, a measure similar to what Zuidema & Westermann (2003) call regularity. The similarity between two sets of attributes, A and B is based on the overlap of the attributes as follows.

$$\text{Alignment}(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{4.1}$$

This measure is known as the Jaccard coefficient (Jaccard, 1901) but most similarity measure for sets would do [e.g. (Sorensen, 1957)]. Some examples:

```
Alignment({a, b, c}, {a, b, c})  = 1
Alignment({a, b, c}, {d, e, f})  = 0
Alignment({a}, {a, b, c})  = 1/3
Alignment({a, b, c}, {a, b, d, e})  = 2/5
```

Both communicative success and lexical alignment are represented as a running average over the last 100 interactions.

4.5.1 Understanding the characteristics of the emergent lexicons

Communicative success and lexical alignment inform us about the effectiveness and the similarity of the emergent lexicons but not about the nature or characteristics of those lexical systems. In a Compositional Guessing Game, as opposed to a Minimal Guessing Game, the diversity in possible emergent lexical systems is much larger. I believe this point is already made to some extent by comparing Figures 4.1 and 4.2 from the introduction. In the most extreme case, and depending on the strategy they follow, agents might invent a word for every possible meaning (subset) for the given set of attributes. For a total of only 20 attributes this would lead to more than one million ($2^{20} = 1048576$) words. Another extreme would to invent only words expressing a single attribute (every meaning is a singleton). In this case the

agents would need only 20 words for 20 attributes, but they would need to utter more words to describe an object. But these are only two extremes, any lexical system in between these extremes could emerge as well.

To gain insight into the nature or the characteristics of the emergent lexicons we measure two aspects at the end of each game.

utterance length: The amount of words in the spoken utterances.

meaning length: The average amount of attributes expressed by the meanings associated with the spoken words of the utterance.

These measures allow us to reveal both aspects of compositionality in the lexicons.

The most important question is then what kind of characteristics we would prefer the emergent lexicons to show if any at all. Are all lexical systems that allows the agents to reach communicative success equally good? This would mean that the two extreme lexical systems sketched above would be equally good (at least if they are aligned enough to reach equal levels of communicative success). One could propose that a lexicon of minimal size that allows the agents to reach communicative success is better than larger sized lexicons. Or one could argue that lexicons that lead to shorter utterances are better. Or maybe a lexicon that strikes a balance between these two constraints? The point I wish to make is that this is not a straightforward question.

In this thesis I follow a usage-based approach by which I mean that an emergent lexicon is better than another when the expressible meanings capture the things the agents have to talk about most. For example assume we need to develop a lexical system that allows us to talk about the animals we find in a zoo. Since we represents objects (animals) as sets of attributes there could be thousands of attributes such as [four_legged], [can_fly], [carnivorous], [animate] and so on. In the fully atomic language sketched above agents would have words for each individual attribute and thus speak sentences such as "flying egg_laying nest_building white_colored (animal)". But in a zoo there tend to be many [flying], [egg_laying], [nest_building] creatures and we call them "birds". I call "bird" usage-based because from a communicative (usage) perspective it makes sense to express that cluster of co-occurring attributes. Rosch *et al.* (1976)[p. 383] explained this as follows:

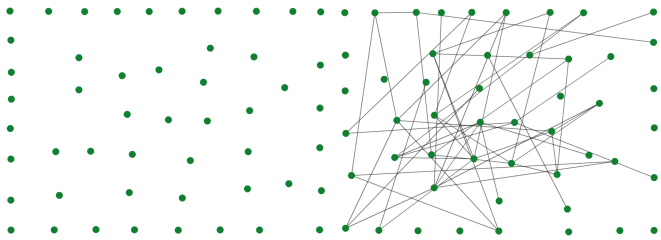
> The world [...] is structured because real-world attributes do not occur independently of each other. Creatures with feathers are more likely also to have wings than creatures with fur, and objects with the visual appearance of chairs are more likely to have functional sit-on-ableness than objects with the appearance

> of cats. That is, combinations of attributes of real objects do not occur uniformly. Some pairs, triples, or ntuples are quite probable, appearing in combination sometimes with one, sometimes another attribute; others are rare; others logically cannot or empirically do not occur.

This means that we want the emergent lexicons to capture co-occurrences of attributes over the objects. If the attributes [flying], [egg_laying] and [nest_building] co-occur frequently we want the agents to capture this. Such a lexicon not only makes sense it also leads to other properties found in human natural languages. For example lexical systems found in the natural world have words of different schematicity Langacker (2000b), they have specific words and general words and many in between. For example, when a dog runs by I can refer to it as a "dog", but also as an more generally as an "animal" or more specifically as a "terrier". This translates to meanings that are subsets or supersets of each other or that partially overlap with each other, just like I chose to depict it in Figure 4.2.

Rosch *et al.* (1976) also points out the importance of basic level categories, which are the categories we prefer to use when referring to a particular object. In the above example "dog" expresses the basic level category, whereas "animal" and "terrier" do not. Although an interesting observation I cannot, with the strategies I present in this thesis, shed much light on this observation and thus will not further delve into that.

If we want the agents to capture co-occurring attributes in the lexicon then we obviously need to supply them with objects that show such co-occurrences. If we want the agents to talk about "birds" we need to present them with birds. In this and the remainder of the thesis I make use of three different types of data-sets from which I draw objects. The first, and it is the only one used in the current chapter, is a context generating algorithm which allows me to influence the statistical distribution of the attributes over the objects in each context. This gives the greatest control over the "world" the agents are faced with. I give more details about this after I introduce the two other types. The second type of data, which I use in chapters 5 and 7 is a machine learning data set taken from the UCI Machine Learning data set repository (Frank & Asuncion, 2010). More details on that in the next chapter. And finally, in chapter 6, I also use embodied data coming from experiments using robots. This data poses particular difficulties stemming from the embodiment.



(a) No edges, no structure (b) 50 edges, medium structure

Figure 4.5: Regulating of the amount of structure in the contexts on the emerging lexicon. Attributes are represented as nodes in a directed graph and attribute nodes that are connected by edges will co-occur. The co-occurrence graph on the left shows no co-occurrence edges and thus represents a highly unstructured world. The co-occurrence graph on the right shows medium structure with the amount of co-occurrence edges equal to the amount of attributes.

4.5.2 Generating structured objects

For the experiments reported in this and most of the following chapter contexts are generated with the possibility to influence the statistical distribution of the attributes over the objects in the context. This makes it possible to create structure in the "world" by introducing co-occurrence relations among attributes.

The context generating algorithm needs to create a set of objects with the constraint that there are no two identical sets (objects) in the context. If there were two identical objects and the speaker would pick one of those as the topic it would be an impossible game to play. The amount of objects in the context is an experimental parameter, by default set to be 5. Each object o_i is a set of attributes drawn from the total set of attributes A. The total set of attributes is fixed from the start of the experiment and the total amount of attributes is again an experimental parameter, by default set to 50. The number of attributes per object is an experimental parameter as well, usually taking a value in between 3 and 10.

But how are the attributes chosen when generating the objects in the context? There are two important steps in explaining this:

1. Before the experiments start we impose co-occurrence relationships unto the attributes through a parameter called the co-occurrence parameter.

2. During the experiment this imposed co-occurrence structure determines

```
Function CreateCo-occurrenceGraph(allAttributes, co-occurrence)
//Distribute weights among the attributes
weight ← 0;
ForEach attribute in allAttributes do
  associate weight with attribute;
  weight ← weight + 1;
End ForEach

//Add co-occurrence edges
i ← 1;
Loop while i <= co-occurrence do
  node1 ← pick attribute from allAttributes taking into account weights;
  node2 ← pick another attribute from allAttributes taking into account weights;
  If node1 and node2 are not yet connected
    then connect node1 and node2;
      i ← i + 1;
  End If
End Loop
```

Algorithm 6: Function CreateCo-occurrenceGraph(allAttributes, co-occurrence) generates a co-occurrence graph given the set of all attributes and a number determining the amount of co-occurrence edges to add.

how objects are composed from attributes.

Starting with (1) imagine the set of all attributes A as nodes in a yet unconnected graph as shown in Figure 4.5a. Each node (or attribute) is given a weight starting from 1 to n, with $n = |A|$. Node 1 receives weight 1, node 2 weight 2, node n weight n. The co-occurrence parameter then determines how many co-occurrence edges are added to the graph. These edges are, however, not added randomly but instead the chance to be connected by an edge depends on the weight of the node. Node n with weight n has n times as much chance to be part of an edge than node 1 has. In short, the weights thus represent probabilities and this probability is distributed linearly over the nodes. Figure 4.5b shows the effect of adding 50 co-occurrence edges. Pseudocode is shown in Algorithm 6

During the experiments (step 2) when generating an object this co-occurrence graph is used to generate each object. The algorithm randomly draws a node (not influenced by edges or weights) but not only adds this node to the object but also all directly connected nodes. It repeats this process until it has reached the amount of attributes per object. Pseudocode is shown in Algorithm 7.

```
Function GenerateObject(co-occurrenceGraph, nr-of-attributes)
object ← empty;
Loop while |object| < nr-of-attributes do
  attribute ← random element from the co-occurrenceGraph not yet in object;
  add attribute to object;
  ForEach connectedAttribute in ConnectedAttributes(attribute,co-occurrenceGraph)
  and while |object| < nr-of-attributes do
    If connectedAttribute ∉ object
    then add connectedAttribute to object;
  End ForEach
End Loop
Return object;
```

Algorithm 7: Function GenerateObject(co-occurrenceGraph, nr-of-attributes) generates an object consisting of n attributes with $n = nr - of - attributes$ according to the co-occurrence graph.

4.6 Experimental results for the Baseline Competitive Strategy

We run a first experiment for a small population of 5 agents, a total of 20 attributes, only 3 attributes per object and 3 objects per context and no co-occurrence edges. Having 3 attributes per object means that a given object can encompass 7 different meanings. For example if the topic consists of three attributes [small yellow triangle], then a single word can express [small], [yellow], [triangle], [small yellow], [small triangle], [yellow triangle], and [small yellow triangle], which are all possible subsets except the empty set which is $2^n - 1$ with n the number of attributes per object.

Figure 4.6 shows us that even under these modest parameter setting the population cannot reach full communicative success. This is quite surprising since the baseline Competitive Strategy is an extension of the Frequency CS Strategy of the previous chapter which was capable of reaching communicative success in seemingly more difficult experimental setups (see for example Figure 3.15 on page 89).

One hypothesis for the failed games, which arose after inspecting those games, is that the problem is to be found in interpretation which, as shown in Algorithm 2, uses only the preferred meaning $f_{meaning}(w)$. As a consequence the listener arrives at a single successful parse which it then checks against the objects in the context. Often this parsed meaning was not compatible with any object in the context. More formal for no object o_i it was true that $\bigcup_{w \in \text{utterance}} f_{meaning}(w) \subseteq o_i$. The interpretation is too strict, often the listener could have inferred the intended object if only he had taken other candidate meanings (beyond $f_{meaning}(w)$) into account.

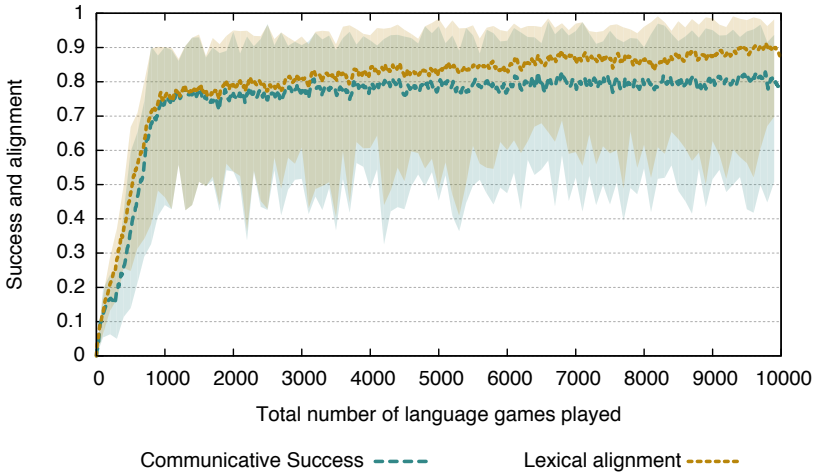


Figure 4.6: Communicative success and lexical alignment for a population of 5 agents following the baseline Competitive Strategy. The population cannot reach either full communicative success nor lexical alignment. Total number of attributes: 20, population size: 5, total games: 4000, 3 attributes per object and no co-occurrence edges, error: 5 and 95 percentile, number of simulations: 20, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$.

To test whether this is indeed the reason let us see whether communicative success improves by extending the interpretation function so that it takes into account all possible interpretations instead of only the "best" one. For this we need a helper function `AllParses(utterance)` which renders every combination of meanings for the words in the utterance. For example, given the utterance "w_1 w_2 w_3" with $f_{meanings}(w_1) = \{\langle a, b\rangle, \langle a, c\rangle, \langle a, f, c\rangle\}$, $f_{meanings}(w_2) = \{\langle a, d\rangle, \langle b, c\rangle\}$ and $f_{meanings}(w_3) = \{\langle e\rangle, \langle h\rangle\}$ there are $3 \times 2 \times 2 = 16$ combinations or possible parsed meanings. Combinations are formed by taking the unions of the respective meanings. For example, taking the first meaning of every word results in the combined meaning (or parse) $\langle a, b, d, e\rangle$. Furthermore `AllParses` returns these parses ordered by the sum of the $f_{freq}(w, m)$ for the used word meaning pairs. For the parse $\langle a, b, d, e\rangle$ this amounts to $f_{freq}(w_1, \langle a, b\rangle) + f_{freq}(w_2, \langle a, d\rangle) + f_{freq}(w_3, \langle e\rangle)$. The end result is an ordered list, based on the summed frequencies, of possible parsed meaning. The new interpretation function is shown in Algorithm 8.

With the extended interpretation function the population can already reach full communicative success (see Figure 4.7). As listeners, the agents

```
———— Function Interpret(utterance, context, agent) ————
interpretations ← ∅;
allParses ← AllParses(utterance,UsageLexicon(agent));
ForEach object in context do
  bestParse ← Find first parse in allParses
                such that parse ⊆ Attributes(object)
  push (object, bestParse) onto interpretations;
End ForEach
Return best of interpretations; //based on sum of frequencies
```

Algorithm 8: Function Interpret(utterance, context, agent) calculates for each object in the context the best possible parse. A parse is a combination of the meanings of the words in the utterance. "Best" is based on the sum of the frequencies of the used word meaning pairs. Finally the function returns the best object/parse pair, again based on the summed frequencies.

are now much more flexible in interpretation and find the intended object with their lexicons. All experimental parameters were kept the same, only interpretation changed.

Still unclear is how the agents describe their topics and what the characteristics of the emergent lexicons are? Since an object is composed of only three attributes, an utterance can consist of no more than three words and a meaning cannot express more than three attributes. The bar charts in Figure 4.8a shows us the percentage of utterances and meanings of lengths 1, 2 and 3 for the final 1000 games. We see that some 90% of all meanings of the usage lexicon of the agents express only one attribute and consequently utterances of three words dominate. Figure 4.8b shows the evolution towards the atomic meanings over the course of the played language games. In the first thousand games meanings expressing one, two and three attributes are floating around but over the course of these first thousand games you can see the alignment dynamics homing in on a highly atomic lexical system. The emergent lexicon thus approximates that of an extremely compositional language with atomic (single attribute) meanings.

4.6.1 The influence of structure in the world

The atomic lexicon the agents arrive at follows from the lack of structure in the world the agents are presented with. Remember there are no co-occurrence edges in the experiments run so far. Without co-occurrence relations it makes little sense to try and express more than one attribute per word meaning. To investigate whether the strategy responds to a change in world structure we adapted the way contexts are generated so that certain attributes are more

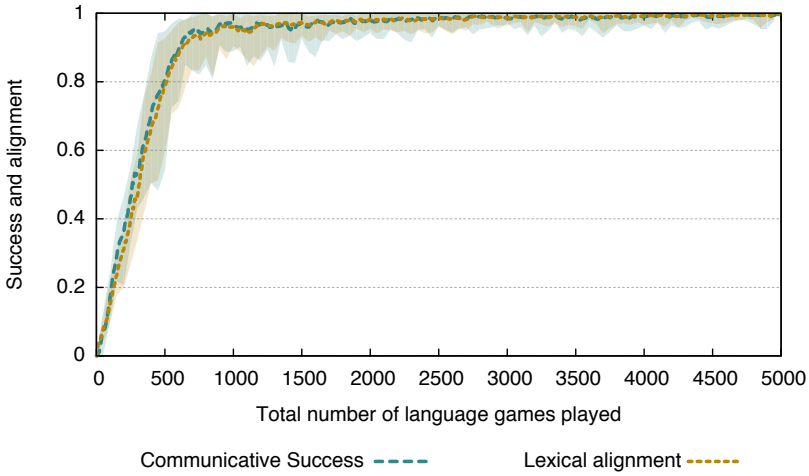
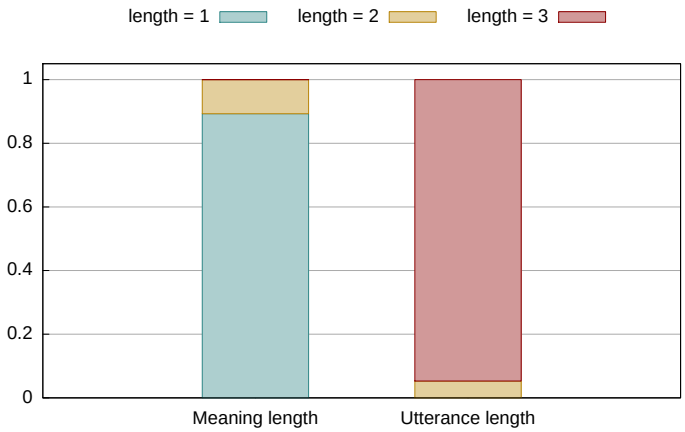


Figure 4.7: Communicative success and lexical alignment for a population of 5 agents following the baseline Competitive Strategy with the extended interpretation function. The population can now reach full communicative success and lexical alignment. Total number of attributes: 20, population size: 5, 3 attributes per object, error: 5 and 95 percentile, number of simulations: 20, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$.

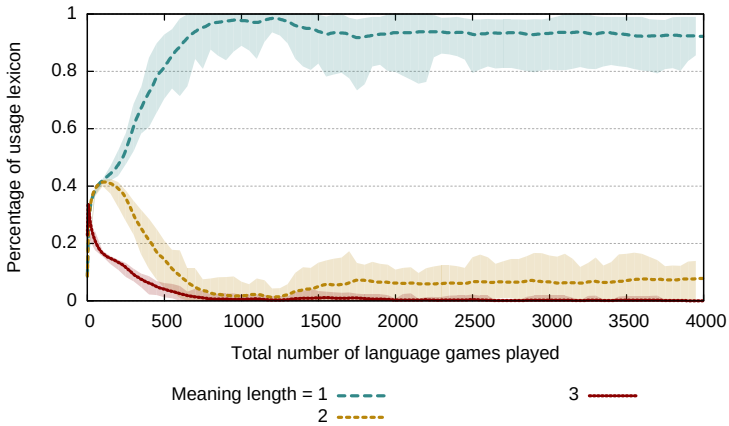
likely to co-occur with other attributes in the same object.

As explained in the Section 4.5, the amount of structure in the world is controlled by the number of edges added to the co-occurrence graph. We introduce for a total of 50 attributes, four conditions with increasing levels of structure or co-occurrence: no edges in condition 1 (highly unstructured world, for example see Figure 4.5a), 20 in condition 2, 50 in condition 3 (for example, see Figure 4.5b), and 100 in condition 4. For each condition a series of 5000 language games with populations of 5 agents was run and the emergent lexicons were compared, as shown in Figure 4.9a.

In condition 1 (no structure), attributes co-occur completely randomly, making any attempt to capture re-occurring patterns in the world useless and this is also what we see emerging in the lexicon. The experiments with the more structured contexts reveal that the agents adapt their lexicons to the changing situation. Even a little bit of structure leads the agents to maintain and use words that capture this structure. This shows that the strategy is not strictly biased towards a certain type of outcome. Finally Figure 4.9b shows that the structured world does make the game harder (as compared to the



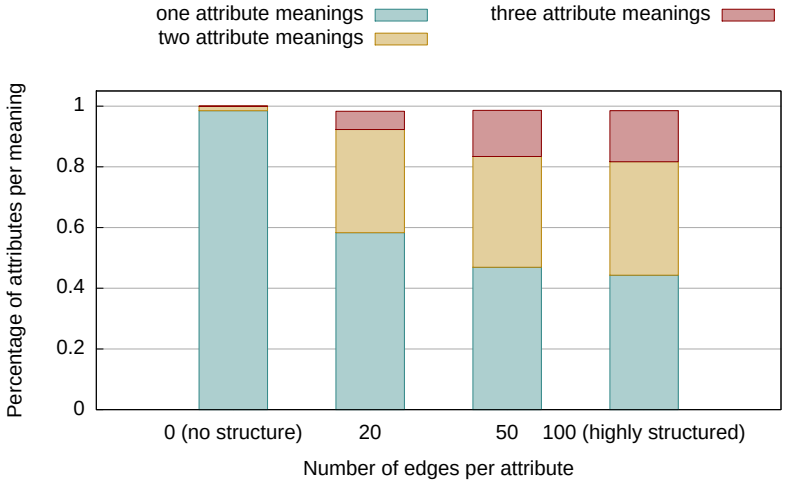
(a) Meaning and utterance length



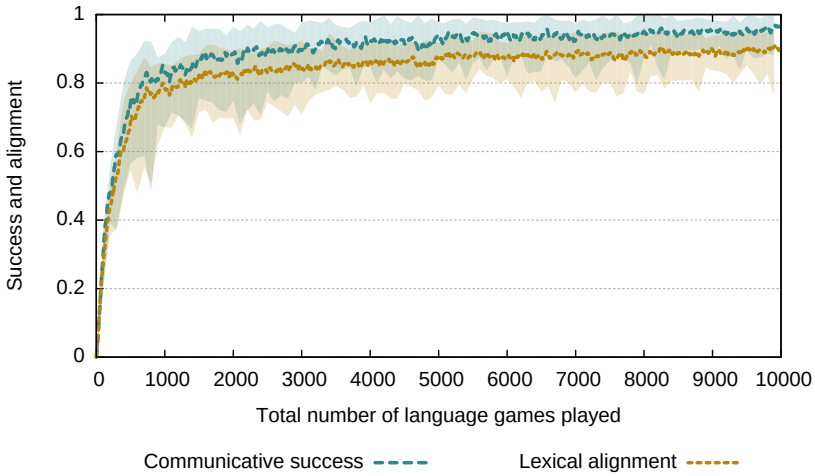
(b) Evolution of meaning length

Figure 4.8: The bar charts in (a) show the stacked percentages of utterances and meanings from length 1, 2 and 3. The agents have adopted an extremely atomic language with meanings expressing only one attribute and consequently utterances of three words. This data is taken from the same experimental runs as those depicted in Figure 4.7. Graph (b) depicts the evolution of the meaning lengths and shows that the agents gradually move towards the atomic language through aligning their lexicons.

.



(a) Impact of structure of context



(b) Communicative success and lexical alignment

Figure 4.9: Effect of the amount of structure in the contexts on the emerging lexicon. (a) The stacked percentages of attributes associated to each word for conditions 1 (no structure) to 4 (highly structured). (b) Communicative success and lexical alignment for experimental condition 4 (100 edges).

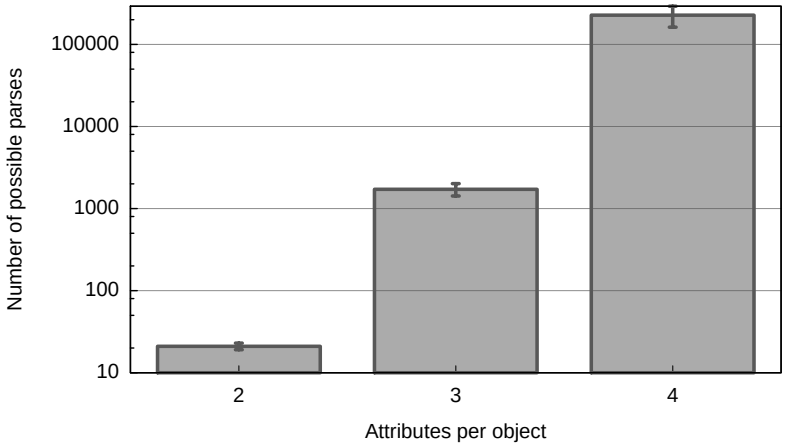


Figure 4.10: Experiments were run with the number of attributes per object ranging from 2 to 4. We show the average number of possible parsed meanings the listener has to check in every game to arrive at a flexible interpretation. Note the log scale on the y axis.

results in Figure 4.7). The agents can still reach high levels of communicative success and alignment yet they do so slower than in an unstructured world. Given the nature of the strategy this makes sense since aligning more complex meanings takes more time.

4.6.2 Issues in scaling

The baseline Competitive Strategy has however one crucial weakness. In the Compositional Guessing Game the number of candidate meanings scales exponentially with the number of attributes per object which is especially problematic since we discovered that the listener needs to be flexible in interpretation and cannot only take into account his reduced usage lexicon. This truly is a combinatorial explosion as shown in Figure 4.10 where for different numbers of attributes per object we show the total number of parses the listener has to check in interpretation (the outcome of function `AllParses(utterance)`). Memory constraints on the available machines did not allow us to scale beyond four attributes per object.

A technique that has often been used in the Compositional Guessing Game is to not express the topic in full but to first find its discriminating features

or attributes and only express those. This leads us to the Discriminative Strategy for the Compositional Guessing Game.

4.7 The Discriminative Competitive Strategy

The Discriminative Strategy for the Compositional Guessing Game differs only in one respect from the baseline Competitive Strategy. Both speaker and listener have the capacity to find a discriminative subset of the attributes of an object (e.g. the topic) with respect to other objects (e.g. the rest of the context). For example given topic [big red box] and another object [tiny red ball], attribute [red] is non-discriminative while any of the others, and thus also their combination, is. The main motivation is that discrimination should substantially improve communicative success and reduce the level of uncertainty. The Discriminative Strategy is still a Competitive Strategy since it also relies on an explicit elimination of competing hypotheses.

4.7.1 Conceptualization and production

Whereas in the baseline strategy the speaker tries to express all attributes of the topic the speaker can also choose to express a subset of those attributes. The process in which the speaker chooses which attributes of the topic to express is called *conceptualization*. In the Compositional Guessing Game conceptualization amounts to a reduction of the attributes of the topic with the aim of improving communicative success.

```
Function Discriminate(topic, context)
validDiscriminations ← ∅;
Loop For i from 1 to Length(Attributes(topic))
  combinations ← CombinationsOfLength(Attributes(topic),i);
  ForEach combination in combinations
    If ∀ o_j ∈ objects \ topic: combination \ Attributes(o_j) ≠ ∅
    then push combination onto validDiscriminations;
  End ForEach;
  If validDiscriminations ≠ ∅
  then return validDiscriminations;
 End Loop;
```

Algorithm 9: Function Discriminate(topic, context) returns the set of minimal discriminations for the topic with regard to the other objects.

In the Discriminative Strategy agents conceptualize an *expressible minimally discriminative* subset of the attributes of the topic. A subset is

```
———————— Function Produce(topic, context, agent) ————————
utterances ← ∅;
lexicon ← UsageLexicon(agent);
minimalDiscriminations ← Discriminate(topic,context);
ForEach candidate in minimalDiscriminations do
  newUtterances ← ProduceCompetitive(candidate,context,agent);
  push newUtterances onto utterances;
End ForEach
utterances ← KeepOnlyShortest(utterances);
Return RandomElement(utterances);
```

Algorithm 10: Function Produce(topic, context, agent) expresses the minimal discrimination of the topic in the context. By using the UsageLexicon production takes into account the frequency and form competition scores of the word meaning pairs in the lexicon.

discriminative in the context when only the topic exhibits all of the features of that set. Discrimination as a form of conceptualization has been used by many others, such as (Steels, 2000; De Beule & De Vylder, 2005; Loetzsch *et al.*, 2008a; Belpaeme, 2002). The discrimination is *minimal* when it does so with the minimal amount of attributes. For example given attributes $\{a, b, c, d\}$ and a context consisting of three objects $o_1 = \{a, b, c\}$, $o_2 = \{a, b, d\}$ and $o_3 = \{b, c, d\}$ then for o_1 there are two discriminative subsets $\{a, c\}$ and $\{a, b, c\}$ with the first being the minimal one. Often there is more than one minimal discriminative subset, in which case the agents prefer the most expressible. Pseudocode for such a discrimination function can be found in Algorithm 9.

The speaker finds the most expressible discrimination by producing an utterance for every minimal discrimination with the function `ProduceCompetitive` from the baseline strategy and keeps only those discriminations that are actually expressible. If multiple discriminations are expressible the speaker prefers the shorter utterances over longer ones, ultimately resorting to a random pick if multiple of those remain. Production for the Discriminative Strategy can be implemented as shown in Algorithm 10.

Interpretation is not modified compared to the flexible version from the Competitive Strategy as explained in Section 4.6, Algorithm 8[1].

[1] A modified strategy was also tested in which the listener interprets not against the full objects in the context but against every minimal discrimination of every object in the context. Although requiring much more processing, this showed no significant improvement.

```
— Function AdoptDiscriminative(utterance, topic, context,agent) —
If unknown forms in utterance; // diagnose
then // repair
  novelForm ← first novel form in utterance;
  interpretedAttributes ← ∪_{w_i ∈ utterance} f_meaning(w_i);
  minimalDiscriminations ← Discriminate(topic, context);
  ForEach discrimination in minimalDiscriminations
  If interpretedAttributes ⊆ discrimination
  then hypothesizedMeaning ← discrimination \  interpretedAttributes;
       addWord(novelForm,hypothesizedMeaning,agent);
  End ForEach
```

Algorithm 11: Function AdoptDiscriminative(utterance, topic, context, agent) diagnoses for unknown word forms. If so he adds hypothesized meanings based on discrimination and the words already known by the agent.

4.7.2 Invention and adoption

Invention remains the same as in the Competitive Strategy (see Algorithm 2) except that the topic is replaced by its chosen discrimination. This leads the speaker to invent words that, on average, express less attributes.

Adoption for the listener takes the discrimination capabilities of the speaker into account as well. To limit the number of initial hypotheses the listener uses the following heuristics:

- He assumes that the speaker expressed a minimal discrimination of the topic and the listener uses the words he knows to guess the conceptualization of the speaker from these minimal discriminations.
- The listener also uses the meanings of the words he already knows to rule out attributes to associate (mutual exclusion).

With these two heuristics the number of initial hypotheses can be brought down considerably. Pseudocode for function `AdoptDiscriminative` is shown in Algorithm 11.

With regard to alignment, steps 1 (form competition) and 2 (frequency updating), remain unchanged from the baseline Competitive Strategy (Section 4.4.4 and Algorithm 5). Only the third step of introducing new competing meanings is updated to incorporate discrimination. As with adoption the listener can only assume that the speaker preferred a minimal discrimination, not which one. He thus first calculates all minimal discriminations and for each of those checks the unexpressed attributes remaining after taking into account the known words (mutual exclusion). He then adds these meanings as potential candidates (see Algorithm 12).

```
Function AlignDiscriminative(words, topic, context, agent)
// Steps (1) and (2) are same as baseline Strategy
If Communication was unsuccessful
then // (3) Update meaning hypotheses
  ForEach <word, meaning> pair in words do
  If ∀ m ∈ f_meanings(word): m ⊈ Attributes(topic)
  then // Add new hypotheses
    expressedAttributes ← ⋃_{word ∈ (words \ word)} f_meaning(word);
    candidateDiscriminations ← BestDiscriminations(topic, expressedAttributes);
    ForEach discrimination in candidateDiscriminations do
      unexpressedAttributes ← discrimination \ expressedAttributes;
      AddWord(word,unexpressedAttributes);
    End ForEach
  End ForEach
```

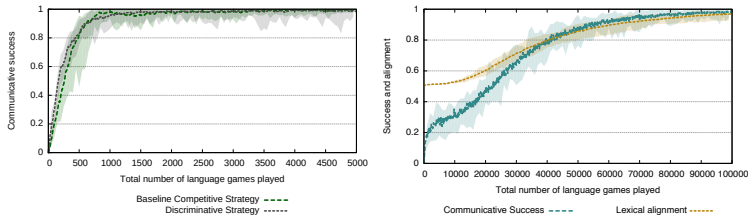
Algorithm 12: Third step in function AlignDiscriminative(words, topic, context, agent). The baseline alignment strategy is modified to take discrimination of the speaker into account. See text for more details.

4.8 Experimental results for the Discriminative Strategy

The Discriminative Strategy can cope with much larger sets of attributes per object as is shown in Figure 4.11b. For a population of 20 agents, a total number of attributes of 50, context size of 4 objects and 10 attributes per object the population reaches full communicative success paired with high levels of alignment. Note that this experimental setup cannot be run with the baseline strategy due to the combinatorially increasing number of interpretations. Figure 4.12 also shows that the number of possible interpretations does not explode with regard to the number of attributes per object. Note that as opposed to Figure 4.10 the y axis has a normal scale this time.

When we look at the emergent lexical system we see that even more than the baseline Strategy the agents have resorted to an extreme atomic language. Word meanings consist of only one attribute and discrimination seems so powerful that single word utterances suffice (see Figure 4.13a). In contrast to the baseline Strategy the atomic meanings are not the emergent result of alignment dynamics but instead are immediately invented and adopted as such, again due to the discrimination capabilities (see Figure 4.13b).

Maybe introducing structure in the world leads to richer and more diverse lexical meanings. With the same methodology as described at the end of section 4.6 we control for the amount of structure in the generated contexts. We took a population of 5 agents, a total of 50 attributes, contexts of 4 objects and each object composed of 10 attributes and changed only the number of edges in the graph that regulates context structure ranging from no structure



(a) Comparison with Baseline Competitive Strategy

(b) Scaling up parameters

Figure 4.11: (a) Shows an experiment with the same parameters as used in the baseline experiments (Figures 4.7 to 4.9). Total number of attributes: 20, attributes per object: 3, population size: 5. (b) Shows that for scaled up parameters the Discriminative Strategy can still reach communicative success and high levels of alignment. Total number of attributes: 50, attributes per object: 10, population size: 20, (a+b) error: 5 and 95 percentile, number of simulations: 20, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$.

at all to an intricate graph of co-occurrence relations.

Figure 4.14 shows the impact of changing the co-occurrence relations. Only when introducing extreme levels of structure do we see a change in the lexicons of the agents. Quite different from the results of the baseline strategy where even a limited amount of structure was immediately reflected in the emergent meanings. The Discriminative Strategy is thus biased towards an atomic lexicon. This means the discrimination component is very successful at reducing the uncertainty. On the other hand it seems too successful in that it also reduces the potential lexicons that can be formed and is not able to take advantage of structured contexts.

4.9 Conclusion

In this chapter I introduced the problems of lexical compositionality and multi-dimensionality in meaning. Meanings are no longer assumed to be atomic or established beforehand but instead need to arise and become aligned through interaction. Agents are no longer uncertain about how to map words to a set of given meanings but rather what the internals of the meanings are, leading to the problem of *meaning uncertainty*.

Since meanings need to be shaped, the range of possible emergent lexical systems is much more open than it was in the Minimal Guessing Game or in the Naming Game. I argued that in order to evaluate and compare

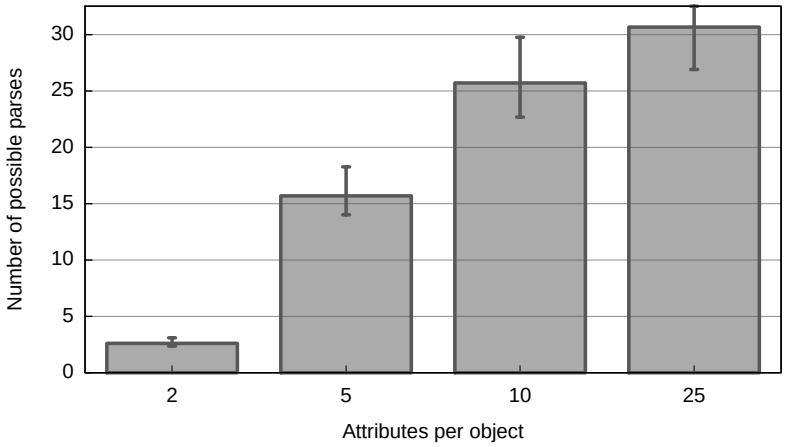


Figure 4.12: Experiments were run with the number of attributes per object ranging from 2 to 25. We show the average number of possible parsed meanings the listener has to check in every game to arrive at a flexible interpretation. Total number of attributes: 100, population size: 5, context size: 5.

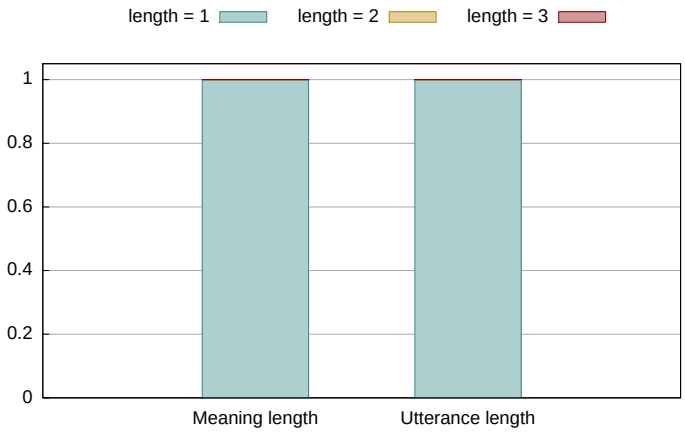
strategies for the problem of meaning uncertainty we should not only look at communicative success and lexical alignment but also at characteristics of the emergent lexicons. Not the usual suspects like minimal lexicon size or minimal utterance length should be aimed for but instead I suggest we should be interested in the following two features:

Schematicity: Do the emergent lexicons show words with both specific and general meanings.

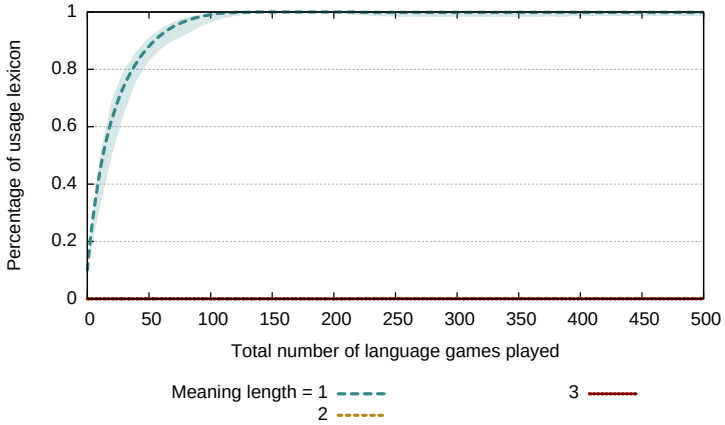
Usage-based: Do the internally structured meanings respond to co-occurrence structure in the objects they need to communicate about.

The intent of this chapter is to investigate to what extent the competitive cross-situational strategies from Chapter 3 can tackle the problem of meaning uncertainty. The Frequency CS and Flexible CS Strategy proved successful in dealing with the mapping uncertainty problem.

The first strategy, the Baseline Competitive Strategy, aims at being the minimal extension from the Frequency CS Strategy that takes into account structured meanings and multi-word utterances. The key point is that it remains true to the basic tenets of competitive learning, which deals with uncertainty by enumerating and subsequently eliminating competitors.



(a) Meaning and utterance length



(b) Evolution of meaning length

Figure 4.13: The bar charts in (a) show the stacked percentages of utterances and meanings of length 1, 2 and 3. Agents adopted an extremely atomic language with meanings expressing only one attribute. In contrast to the baseline Strategy utterances consist of only a single word. Graph (b) shows the evolution of the meaning lengths zoomed in on the first 500 games and shows that the atomic meanings are not the results of a gradual emergent process but instead directly derive from the discrimination capabilities.

.

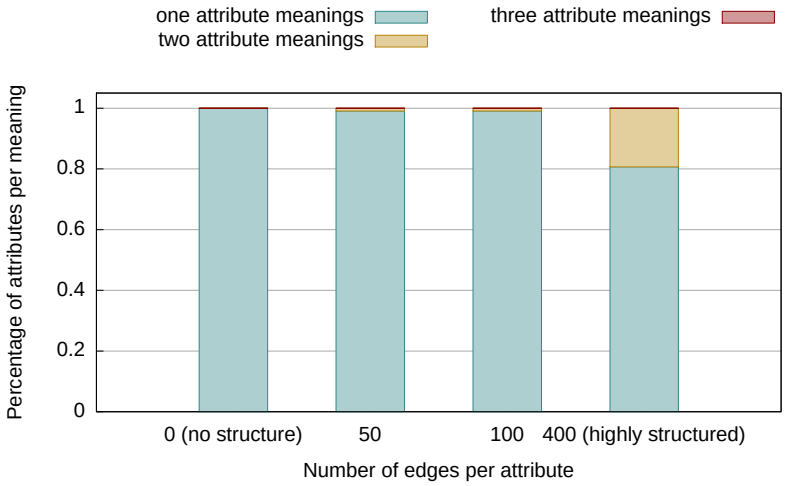


Figure 4.14: Influence of structure in the contexts on the formed lexicons. With the Discriminative Strategy the impact of a structured world is considerably less than with the baseline Strategy.

In the first implementation agents did not reach full communicative success and lexical alignment even under conditions of low uncertainty. The reason, it turned out, is that the default interpretation function only looked at the currently preferred meaning $f_{meaning}(w)$ for each spoken word w. Although this kept interpretation fast it also made it too strict.

Opening interpretation to take into account more meanings than only the currently preferred meaning turned out to be a Pandoraś box. Populations could now reach full communicative success and full lexical alignment in low uncertainty conditions, proving that indeed the strictness of interpretation was the root of the problem. Furthermore in these low certainty conditions the emergent lexicons even showed usage-based properties and schematicity in the established meanings. But the more flexible interpretation also prohibited the strategy from scaling to high uncertainty conditions.

A common approach to limit the number of candidate meanings is to introduce word learning constraints or biases. There is clear evidence that children use behavioral cues to establish joint attention and determine the speaker's attentional focus (Baldwin, 1991; Hammer & Diesendruck, 2005; Blair & Somerville, 2009; Nappa *et al.*, 2009; Graham *et al.*, 2010). Tomasello & Carpenter (2005) even speak of *intention reading*, stressing a child's extraordinary capacity to recognize others intentions. Other research has focused on

	Naming Game	Minimal GG	Compositional GG
meaning complexity	atomic	atomic	structured (set)
utterance complexity	single word	single word	multi word
word form competition	yes	yes	yes
mapping uncertainty	no	yes	no
meaning uncertainty	no	no	yes

Table 4.1: Overview of Naming Game and Minimal Guessing Game. The Minimal Guessing Game differs only in one respect from the Naming Game in that it evokes mapping uncertainty.

the types of information children may rely on to form categories of objects (Madole & Oakes, 1999; Rakison & Oakes, 2003). Many of these constraints are taken for granted in the proposed strategies, including those in the next chapter. For example during learning (adoption and invention), the constraint of mutual exclusion is taken into account and to a certain extent those of attentional focus since a listener only learns after extralinguistic feedback. Even with those constraints in place the uncertainty is too large for the baseline Competitive Strategy.

The root of the problem lies in the enumeration of hypotheses which, if meanings are sets of attributes, scales exponential in the number of attributes per object. This analysis is acknowledged by the Discriminative Strategy which is the Baseline Competitive Strategy combined with an extra pre-processing "cognitive" ability aimed at reducing the number of candidate hypotheses. In the Discriminative Strategy this pre-processor is a capacity to find, given a topic and a context, a minimal set of attributes in the topic which are only part of the topic and are thus discriminative for the topic. Expressing only those discriminative attributes should lead to communicative success and at the same time severely limit the number of attributes to consider as part of the initial meaning competitors. In short, instead of all possible subsets only minimally discriminative subsets become viable candidates as initial meaning hypotheses.

The results from adding this extra discriminative component were at first very positive. Populations could now bootstrap a shared lexicon even when the population size and attributes were scaled so that high levels of uncertainty could be expected. Since everything remains the same between the two strategies except for the added discrimination component the answer must lie there. And indeed, upon inspecting the emergent lexicons we see that they tend to be highly atomic with words expressing only a single attribute. Further investigation into how these meanings emerged showed that it was not the result of a gradual dynamics but that these meanings were

invented and adopted as such. This shows that the discrimination component has effectively reduced the problem of meaning uncertainty to a problem of mapping uncertainty. This explains why the competitive cross-situational learning strategy works because it is solving a much simpler problem.

This reductionist approach comes at a steep price because by reducing the problem to a mapping problem it at the same time diminishes the range of potential emergent lexical systems. Fully atomic lexicons do not show schematicity as all words are maximally general. All word meanings are also disjoint, with no meaning sharing the least bit of similarity with another one. Most importantly the strategy is not responsive to usage-based pressure. If we introduce re-occurring patterns in the world then the lexical systems no longer capture these patterns.

In Chapter 6 we again compare the Discriminative Strategy with the alternative strategy introduced in the next chapter and there we will uncover another weakness in such a strategy. For example, when the attribute sets of the agents are not entirely the same (e.g. because of noise or different perceptions) then discrimination might magnify this and severely undermine the goal of reaching high levels communicative success.

In conclusion competitive strategies do not truly cope or convincingly solve the problem of meaning uncertainty. In the next chapter I introduce a different approach to cross-situational learning which is not competitive but stresses flexible re-use of existing meanings and an internally adaptive meaning representation.

Part III

Uncertainty in lexical systems: adaptive strategies

> Our minds delight in the discovery of resemblances, near and remote, obvious and obscure, and are always ready to make them the foundation of an association that involves the addition of a new use to an old name.
>
> William Dwight Whitney (1875, p. 86)

Two views of word learning

In the previous part we introduced competitive cross-situational strategies and showed how they can successfully tackle the problem of mapping uncertainty (Chapter 3) but are less convincing when confronted with the problem of meaning uncertainty (Chapter 4).

Competitive strategies are compatible with a view of word learning as *mapping*. A word learner has to map a set of forms onto a set of pre-established concepts (Bloom, 2000). The assumption is that learners have access to a number of potential meanings and need to choose (or guess) the correct one. For example Bloom (2000) writes "what goes on in word learning is establishing a correspondence between the symbols of a natural language and concepts that exist prior to, and independently of, the acquisition of that language". Siskind (1996) very clearly defines the lexical acquisition task as such a mapping problem. In Chapter **??** we have seen that competitive cross-situational learning strategies, especially those that allow the agents to track the cross-situational statistics through a scoring mechanism, are very well suited for this problem.

In Chapter 4 we introduced a more difficult interpretation of the word learning task where the problem is no longer that of a mapping task but that of constructing multi-dimensional word meanings. A straightforward application of a competitive learning strategy to this more difficult problem of meaning uncertainty did not scale well as the meaning space introduced by a Compositional Guessing Game was too vast. The core of the problem lies in the large number of possible competing meanings, a problem recognized already a long time ago. Since then a large body of research has focused on finding word learning constraints (or biases) that guide a learner towards the right mapping and that significantly reduce the number of initial competing mappings (see for example (Gleitman, 1990; Bloom, 2001; Markman, 1992)).

Although I do not argue against such constraints, in fact I implement some of those myself in this chapter, they are proposed to solve a problem which still stems from viewing word learning as mapping.

The Discriminative Competitive Strategy presented in the previous chapter can also be seen as embodying that approach to word learning. It is still competitive in that it enumerates competing hypotheses and it still sees word learning fundamentally as mapping. By adding extra cognitive capabilities or biases which apriori reduce the hypothesis space it does allow agents to bootstrap a communication system in larger spaces. The Discriminative Strategy added a discrimination component to all agents which, as we have seen, was so successful in reducing the number of hypotheses that it also heavily constrained the possible emergent lexical systems. Most importantly the emergent lexical systems did not exhibit a wide range of meanings ranging from specific to general meanings (only atomic, general meanings were found) and the lexical systems failed at capturing re-occurring patterns in the objects they needed to communicate about. In the following chapter we highlight another weakness of relying so heavily on discrimination.

In short we have seen in Chapter 3 that competitive approaches are sufficient when meanings are taken to be atomic. Extending the competitive approach to compositional meanings did not lead to a convincing outcome. This is why often competitive cross-situational learning goes together with a set of constraints or biases: word learners use constraints to make a limited list of initial mappings and rule out all except one hypothesis later on.

There is a growing body of researchers challenging the mapping view of word learning and suggesting that the meanings of words might not yet be readily available to the child. Bowerman & Choi (2001) [p. 476] disagree with a view in which "The child is characterized as needing to *identify* the concept among a set of plausible possibilities." Instead they envision the child as constructing and gradually shaping word meanings. In the same volume Tomasello (2001) describes the competitive approach as follows:

> One approach to the problem of referential indeterminacy in the study of lexical acquisition is the so-called "constraints" approach. In this view a learner, [...], attempts to acquire a new word by: (1) creating a list of possible hypotheses about how the new word "maps" onto the real world, and (2) eliminating incorrect hypotheses in a semi-scientific manner.

His own view of word learning lies in the line of that of Bowerman & Choi (2001). Tomasello (1999) puts it as follows: The hypothesis is that "... the use of words in repeated discourse interactions in which different perspectives are explicitly contrasted and shared, provide the raw material out of which the

children of all cultures *construct* the flexible and multi-perspectival – perhaps even dialogical – cognitive representations that give human cognition much of its awesome and unique power" (Tomasello, 1999, p. 163, my emphasis).

Although in this view learners are equally or even more uncertain about meanings of novel words, the uncertainty is of a different nature. Children cannot have at hand all the concepts and perspectives that are embodied in the words of the language they are learning – they have to construct them over time through language use. "For example, many young children overextend words such as *dog* to cover all four-legged furry animals. One way they home in on the adult extension of this word is by hearing many four-legged furry animals called by other names such as *horse* and *cow*" (Tomasello, 2003, pp 73–74). Moreover, the enormous diversity found in human natural languages (Haspelmath *et al.*, 2005; Levinson, 2001) and the subtleties in word use (Fillmore, 1977) suggest that language learners can make few apriori assumptions and even if they could, they still face a towering uncertainty in identifying the more subtle aspects of word meaning and use.

Chapter 5

Adaptive strategies for Compositional Guessing Games

In the previous chapter we presented two competitive strategies inspired by the cross-situational strategies of Chapter 3. The strategies were proposed to tackle the problems introduced in the Compositional Guessing Game where meanings are composed of multiple attributes (multi-dimensionality) and utterances of multiple words (compositionality). What competitive strategies have in common is that they represent uncertainty by enumerating different competing meanings. The list of competing meanings is gradually reduced by new exposures until a single meaning remains.

In the current chapter a class of strategies is introduced which pays considerable attention to the adaptive nature of meaning. In this strategy agents avoid enumeration of meaning competitors and instead gradually *shape* the meanings of their words. The strategy is based on *re-use*, allowing agents to construct multi-word utterances in which word meanings are not (yet) fully compatible with the topic. Word meanings are adaptive and changed based on their use in previous language games. The details of these strategies are given in Sections 5.1 and 5.3.

The language game setup is exactly the same as explained in Section 4.2 (the Compositional Guessing Game). In short the compositional guessing game is a routinized interaction in which a speaker tries, using language, to draw the attention of a listener to a particular object in a shared scene. The speaker and listener give each other feedback as to whether the interaction was successful and point to the intended object in cases of failure. This allows the population, over the course of many interactions, to self-organize a language for communicating about objects. Objects in each scene are composed of attributes and meanings of words can likewise comprise multiple attributes. Note that agents are implemented such that they do not have access to internal

representations of other agents – there is no meaning transfer, telepathy or central control.

5.1 The Baseline Adaptive Strategy for the Compositional Guessing Game

The main weakness of the competitive strategies lies in the enumeration of competing hypotheses. Constraints-based approaches such as the Discriminative Strategy essentially recognize this to be the core weakness and try to minimize the initial set of competitors. Adaptive strategies take a more radical stance and do not maintain any enumeration or listing of competing hypotheses, but instead each word has only a single meaning hypothesis associated.

This is quite a bold feature of adaptive strategies because it essentially means the representational power and memory capacity needed by an adaptive strategy is much less extensive than that of the competitive strategies. The memory is thus more like a Naming Game memory where each word just has a single associated meaning, although now the meaning is a set. In the Baseline Adaptive Strategy this is exactly the case, each word is associated to exactly one meaning, which is a set of attributes.

Obviously maintaining only a single hypothesis alone isn't going to solve the problem of uncertainty. It does of course "solve" the problem of an explosion of hypotheses in a rather crude manner. The main motivation for this bold choice lies however in the explicit goal not to see word learning as a mapping task. With adaptive strategies we want to investigate the power of constructing and adapting word meanings instead of eliminating competitors.

Maintaining only a single meaning per word is only the first out of three features that a strategy needs to be called adaptive. The second feature relates to how the meanings can be used during processing which, as will be explained in Section 5.1.1, needs to be more flexible. The third core feature of an adaptive strategy pertains to alignment in which speaker and listener internally adapt the meaning of each spoken word.

The Baseline Adaptive Strategy operationalizes these three core features in the most minimal and basic way. Not only can it inform us about the strength of these features when implemented in the most basic way but it also allows us to better evaluate the impact of the second adaptive strategy introduced later in Section 5.3.

With adaptive strategies there is no longer need for the function $f_{meanings}(w)$ because there is only one meaning per word which is returned by function $f_{meaning}(w)$. The notion of a "preferred" meaning for given word w is also no

longer needed as there is only a single meaning per word.

Just like competitive models need to be extended with a word from competition damping mechanism so do adaptive strategies. Just as we have done in the previous chapter each word meaning association also maintains a score $f_{score}(w)$, used for damping form competition. Unless otherwise stated the initial value of $f_{score}(w)$ is 1. There is no score $f_{freq}(w, m)$ or any other score like was required for the Cross-situational and Competitive strategies.

The concept of a usage lexicon (explained in Section 4.4.1) remains. Although, where it previously reduced the lexicon in two ways (removal of competing meaning and form competitors), it now only removes form competitors as there are no competing meanings. We only introduce the usage lexicon in order to simplify the production algorithm. Calculating the usage lexicon does not truly remove these competitors from the complete lexicon, it only generates a reduced version which can be used in processing.

5.1.1 Flexible processing: production and interpretation

Adaptive strategies put re-use of existing words at the center of linguistic processing. Agents re-use words even when some characteristics of the meaning mismatch with the topic. For example if the topic is composed of attributes [green small box] an agent is still allowed to use the word "dobido" even if it has associated meaning [green tall box] with a mismatch on attribute [tall]. To calculate the overlap between words and objects agents can rank their words for a given set of attributes by employing a *similarity function* called `Overlap`. The importance of similarity for cognitive processing has been underlined by a large number of cognitive scientists (Gärdenfors, 2000; Gentner & Rattermann, 1991)

Meanings and objects share the same representation, they are both sets of attributes. The similarity measure between meanings and objects is defined as follows:

$$\texttt{Overlap}(M, O) = \frac{|M \cap O| - |M \setminus O|}{|O|} \tag{5.1}$$

Note that this is not the same measure as was used for measuring lexical alignment in the previous chapter. `Overlap` returns a real value in $[-1, 1]$[1] and has the following properties:

- If $M = O$ then `Overlap(M,O)` $= 1$
- If $M \cap O = \emptyset$ then `Overlap(M,O)` $= -1$

[1]This only holds if $|O| \geq |M|$, which is always the case in the reported experiments.

```
———————————— Function Produce(topic, agent) ————————————
lexicon ← UsageLexicon(agent);
utterance ← nil;
Loop
   bestNewWord ← argmax_{w ∈ lexicon} Overlap(w ∪ utterance, topic);
   newSimilarity ← Overlap(utterance ∪ bestNewWord, topic);
   If newSimilarity > Overlap(utterance, topic) >= 0.0
   then utterance ← utterance ∪ bestNewWord;
   else Return utterance;
End Loop;
———————————————————— End Produce ————————————————————
```

Algorithm 13: Function Produce(topic, agent) incrementally constructs an utterance by adding a new word that most increases the overlap with the attributes of the topic.

- If $M \subset O$ then `Overlap(M,O)` > 0. This feature will be used in processing.
- If M shares more attributes with O than it differs (if $|M \cap O| > |M \setminus O|$) then `Overlap(M,O)` > 0

Essentially matching attributes between M and O pull the similarity to 1 while mismatching attributes pull towards -1. Some examples:

$$\texttt{Overlap(\{a, b, c\}, \{a, b, c\})} = \frac{3-0}{3} = 1$$
$$\texttt{Overlap(\{a, b, c\}, \{d, e, f\})} = \frac{0-3}{3} = -1$$
$$\texttt{Overlap(\{a\}, \{a, b, c\})} = \frac{1-0}{3} = \frac{1}{3}$$
$$\texttt{Overlap(\{a, b, c\}, \{a, b, d, e\})} = \frac{2-1}{4} = \frac{1}{4}$$

Also note that `Overlap` is not commutative, `Overlap(A,B)` need not be equal to `Overlap(B,A)`. This non-commutativity captures the assymmetry in matching meanings with objects as opposed to objects with meanings. The more attributes of a meaning are missing from the intended topic the less appropriate the word becomes. But the opposite is not true, a word can still be very appropriate even though the object contains many attributes not expressed by the meaning. Indeed an agent need not necessarily express all attributes.

Production and interpretation both rely heavily on this similarity measure. The speaker constructs his utterance quite like the agents from the Baseline Competitive Strategy. He tries to express as many attributes of the topic by incrementally adding a new word such that the overlap with the topic is most improved (see Algorithm 13). The main difference is that the overlap is no longer determined by a subset relation but the similarity measure outlined above. A minimum similarity of 0 is required at all times between the

utterance and the topic, which simply means that the constructed utterance should at least be more similar than dissimilar to the topic. If the agent cannot find an utterance (or even a single word) with positive similarity it simply means the lexicon is not expressive enough for the given topic.

Since there is only one meaning associated to each word there is only one possible parse for an utterance U, namely $f_{meaning}(U) = \bigcup_{w \in U} f_{meaning}(w)$. The listener points to the object in context C that maximizes the overlap with the parsed meaning: topic $\leftarrow \underset{o \in C}{\operatorname{argmax}}\, \texttt{Overlap}(f_{meaning}(U), o)$.

Production and interpretation might seem to differ only superficially from Competitive strategies but that is certainly not the case. The use of the similarity measure allows the agents much more flexibility in both production and parsing. The agents can re-use words with meanings that are not subsets of the topic. Adaptive strategies thus allow a mismatch between the intended or interpreted meaning and the topic. Not only does this radically improve processing but the information about matching and mismatching attributes becomes pivotal during alignment.

5.1.2 Invention and adoption

Invention works almost exactly the same as in the Baseline Competitive Strategy (see Chapter 4, Section 4.4.3, Algorithm 3, page 110). Through re-entrance the speaker interprets his constructed utterance and invents a new word when the interpretation does not lead him to the topic. The initial meaning consists of all unexpressed attributes of the topic and thus also implements the constraint of mutual exclusivity (Markman *et al.*, 2003).

There is one difference with invention compared to the Competitive variant which follows from the fact that agents can internally alter (see 5.1.3) their meanings and can measure similarity. The combination of these two capabilities allows a speaker to determine the chance that only a small change to one of the existing words might solve the situation instead of inventing a completely new word. Therefore high similarity between the meaning of the utterance and the topic translates to a lower likelihood of introducing a new word. In pseudocode the invention process can be operationalised as follows:

When the listener encounters one or more novel words in the utterance he needs a way to associate an initial representation of meaning with the novel forms. Due to the internally adaptive nature of the meaning representation this initial representation need not be final. In many cases the meaning hasn't even become conventionalized throughout the population which means there is no fixed intended meaning yet. Even if there was already a fixed meaning then the uncertainty upon first exposure scales exponentially with the number

```
──────── Function Invent(utterance, topic, context, agent) ────────
// Diagnosing the utterance through re-entrance
interpretedObject ← Interpret(utterance,context,agent);
If interpretedObject = topic
then Return utterance;
else // Extend lexicon
  interpretedMeaning ← ⋃_{w ∈ utterance} f_meaning(w);
  topicSimilarity ← Overlap(interpretedMeaning, topic);
  randomNr ← Random(0, 1); // A random number between 0 and 1
  If randomNr > topicSimilarity
  then // The higher similarity the less likely to invent
    initialMeaning ← Attributes(topic) \ expressedMeaning;
    addNewWord(randomForm,initialMeaning,agent);
    Produce(topic, context, agent); // Try again
  End If
End If
```

Algorithm 14: Function Invent(utterance, topic, context, agent) first diagnoses whether a new word should be added. If so it creates a new word with as meaning the unexpressed attributes of the topic.

of attributes per object as seen in the previous chapter.

The listener first interprets the words he knows and tries to finish the game without adopting the novel forms. In case there are multiple unknown forms the listener only learns a meaning for the first one he encounters as was also the case for the competitive strategies discussed in the previous chapter. The reasoning stays the same as it was there. At the end of the game, when the listener has knowledge of the topic, all unexpressed features are associated to the chosen novel form (see Figure 5.1). Excluding the features of the already known words is the only constraint shaping the initial representation.

5.1.3 Alignment

Alignment is where the bulk of the work in Adaptive Strategies takes place. Alignment in Adaptive Strategies consists, like it did in Competitive Strategies, of three steps. First the agent updates the form competition scores $f_{score}(w)$ using interpolating lateral inhibition. This is thus the same updating mechanism as we have been using in all strategies of the previous two chapters and can be found in Chapter 2, Section 2.4, page 51.

The second alignment step is that of adjusting or eroding the meanings of the used words. For each interpreted word the agent calculates the mismatching attributes, which are the attributes that did not match with the topic

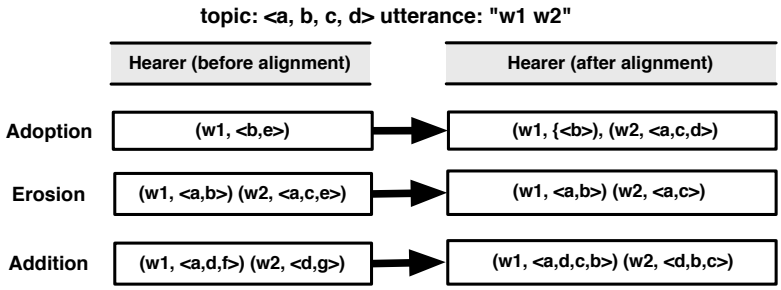


Figure 5.1: Alignment scenarios for the Baseline Adaptive Strategy. The topic consists of attributes a,b,c,d and the utterance is "w1 w2". In adoption the listener adds the new word with all the unexpressed attributes of the topic. Erosion is the removal of attributes not compatible with the topic. In addition all unexpressed attributes of the topic are added. Addition only takes place in failed games without novel word forms.

during interpretation.

$$\text{mismatch(w)} \leftarrow f_{meaning}(w) \setminus \text{topic}$$

The agent removes all mismatched attributes for each used word. For example topic $\langle$a,b,f$\rangle$ and utterance "$w_1 w_2$" with lexical entries (w_1, $\langle$a,b,c$\rangle$) and (w_2, $\langle$a,d,f$\rangle$) results in the lexical entries (w_1, $\langle$a,b$\rangle$) and (w_2, $\langle$a,f$\rangle$). This process can thus reduce the number of attributes from the meaning and is hence called *erosion*. In all three cases in Figure 5.1 meanings are eroded. It might happen that a word meaning loses all associated attributes through erosion. In this case the word is removed from the agents memory.

In the third and final alignment step the listener extends the meanings of the interpreted words if necessary. When the listener knew all the words in the utterance but yet the game was a failure there is a large chance that there is a severe alignment mismatch between speaker and listener. The listener cannot know whether it is only one particular word or multiple words that are misaligned and thus extends the meanings of all the words in the utterance by adding all the unexpressed attributes of the topic (see Figure 5.1). Most likely superfluous attributes are added in this addition step so the listener counts on erosion processes in later games to again reduce the extended meaning. Example code can be found in Algorithm 15.

The combination of erosion and extension processes leads to a dynamics where meanings are internally *shaped* over time towards the emergent conventionalized meaning. In the Baseline Adaptive Strategy this shaping is

```
———————— Function Align(usedWords, topic, agent) ————————
ForEach word-meaning pair <w,m> in usedWords do
  // (1) Dampen word form competition
  f_score(w,m) ← f_score(w,m) + δ_inc(1 − f_score(w,m));
  ForEach word-meaning pair <cw, cm> in FormCompetitors(w, m) do
      f_score(cw,cm) ← f_score(cw,cm) − δ_inh f_score(cw,cm);
  End ForEach

  // (2) Erode meanings
  m ← m ∩ topic;
End ForEach

If Communication was unsuccessful
then // (3) Update meaning hypotheses
  ForEach (word, meaning) pair in usedWords do
    expressedAttributes ← f_meaning(usedWords \ word);
    unexpressedAttributes ← Attributes(topic) \ expressedAttributes;
    AddAttributes(word,unexpressedAttributes);
  End ForEach
```

Algorithm 15: Function Align(usedWords, topic, agent). See text for details.

unrefined in that it only allows the removal or addition of attributes. Later strategies will have more subtle and gradual shaping dynamics.

Zuidema & Westermann (2003) presented a model very similar to the Baseline Adaptive Strategy in terms of representation and processing. Like the Baseline Adaptive strategy agents could employ a similarity measure in a compositional meaning space and did not enumerate hypotheses like a competitive model would do. The biggest difference lies in learning and alignment. Essentially the agents do not align at the end of a language game. More specifically the agents do not employ the mismatch information that could be obtained from processing, in fact they do not use any information that could be obtained from the current game at all. Instead random mutations to the lexicon are tried out and a fitness in terms of overall communicative success is calculated. Once a mutation is found that would improve overall communicative success of the agent the change is accepted. Because the agents lack a directed alignment step the overall lexical alignment (called regularity by them) remains significantly lower although it does increase since all agents are moving towards a lexicon that allows them to communicate better.

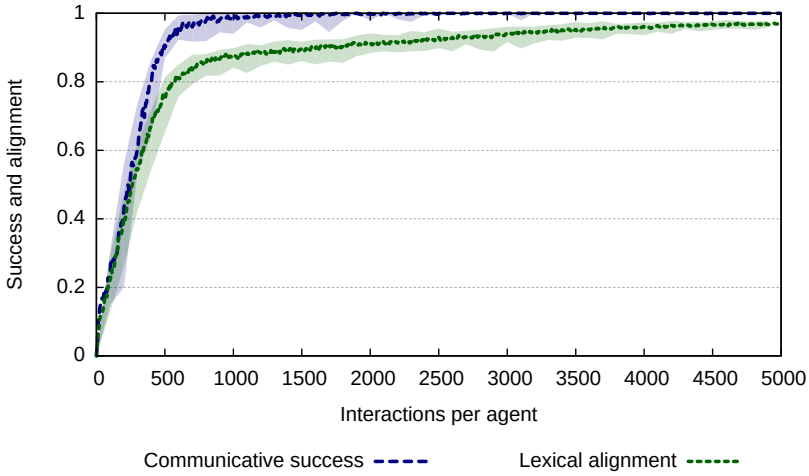


Figure 5.2: Communicative success and lexical alignment for a population of 10 agents following the Baseline Adaptive Strategy. The population reaches full communicative success and keeps gradually improving lexical alignment. Total number of attributes: 100, number of attributes per object: 10, co-occurrence edges: 100, population size: 10, total games: 25000, error: 5 to 95 percentile, number of simulations: 20, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$.

5.2 Experimental results for the Baseline Adaptive Strategy

Since the language game has not changed from the previous chapter the measures communicative success and lexical alignment can be reused (see Section 4.5). Unless otherwise stated contexts are always structured by the method explained in Section 4.5, Chapter 4, page 114. By default the number of co-occurrence edges will be equal to the total number of attributes. For example in most experimental runs a total of 100 attributes is present and thus also a total of 100 co-occurrence edges.

For a first experimental run with a population of 10 agents, a total of 100 attributes with 10 attributes per object and a context size of 5 we see in Figure 5.2 that the population reaches full communicative success and high levels of alignment. What is interesting is that high levels of communicative success are reached very early on and that lexical alignment keeps improving even though full communicative success has been reached. This are the shaping dynamics

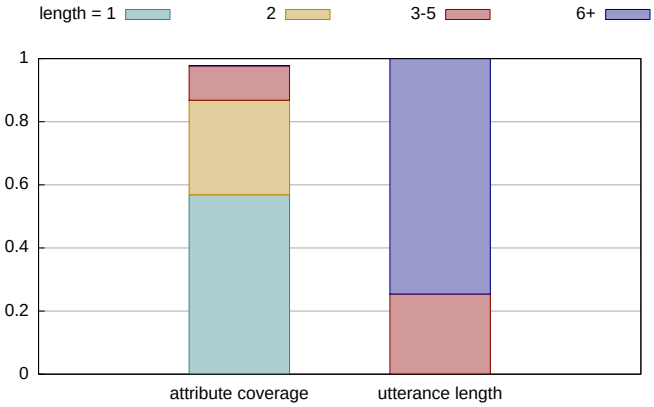
at work, bit by bit improving the conventionalized meanings. Note that the Baseline Competitive Strategy cannot operate under these parameters.

Figure 5.3 gives more information about the characteristics and use of the emergent lexicon. The left bar in part (a) shows for all spoken words the percentage that express a certain amount of attributes (1, 2, 3-5, 6+). For example, slightly less than 60% of all spoken words express only a single attribute, 32% express 2 attributes and the remaining words express 3 to 5 attributes. This result shows that the lexicon emerging from the Baseline Adaptive Strategy has captured some of the structure in the world and did not result in a purely atomic lexicon like the Discriminative Strategy of last chapter did. The bar at the right-hand side shows the percentage of utterances that consist of only a certain amount of words. Utterances tend to be long with some 70% of all utterances consisting of 6 or more words. This is not surprising since the agents try to exhaustively express all attributes of the topic.

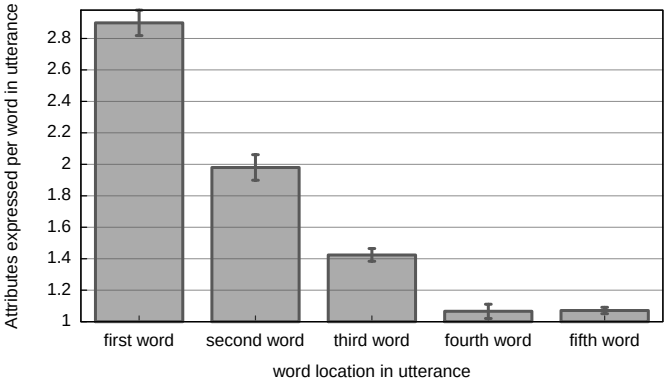
Part (b) of Figure 5.3 shows in more depth the internal structure of the constructed utterances. Each bar corresponds to a word in an utterance, the first bar to the first words, the second bar to the second and so on. The bar itself depicts the average number of attributes expressed by the word in that position. On average the first word in an utterance expresses almost three attributes out of 10, the second word expresses on average two more. This means that the first two words express on average half of the topic. The remaining words then fill in the gaps that are left. This ordering is a direct consequence of the production algorithm which first favours most similar words which tend to be multi-attribute words. This ordering is thus not an emergent effect but what is interesting is that it shows that two types or classes of words have emerged. There are cluster words which express multiple attributes and they always come first and then there are atomic category words which express only a single dimension. Most of these category words were once more holistic cluster words but through erosion processes ended up as category words. In the next chapter we further investigate and visualise these erosion processes (e.g. Section 6.2.1).

5.2.1 Impact of co-occurrence structure

Let us now investigate the performance of the strategy and the characteristics of the emergent lexicons when manipulating context structure, population size, and the number of attributes per object. Context structure is manipulated in the same manner as in the previous chapter, by changing the number of edges in the co-occurrence graph. When attributes co-occur in objects it should, in the long term, be beneficial to express those attributes with one



(a)



(b)

Figure 5.3: Characteristics of the emergent lexicon. (a) The left bar shows the percentages of words that express 1, 2, 3-5, or over 6 attributes. The right bar shows the percentage of utterances that consist of 3-5 or over 6 words. (b) For the first five words of each utterance a bar represents the amount of attributes expressed by the word in that location. The first words express clusters of attributes while remaining words fill in the remaining gaps. (a and b) Total number of attributes: 100, attributes per object: 10, population size: 10, total games: 50000, number of simulations: 4, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$.

word. Co-occurrence of attributes should thus lead to lexicons with meanings expressing more attributes on average. Another question is whether the structure in the world helps the agents in bootstrapping their lexicons, a feature we did not find in the competitive strategies.

For the same experimental parameters as before a condition was added in which the objects in the context exhibit no structure (zero co-occurrence edges). Figure 5.4a shows that also in the non-structured condition the population can reach full communicative success although structured contexts help in bootstrapping the lexicon. The competitive strategies did not show this benefit of structure.

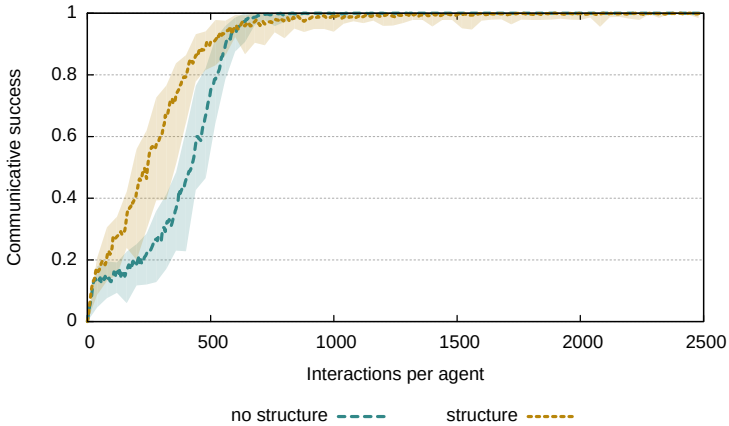
The explanation can be found in Figure 5.4b. Each bar shows the percentage of word meanings that express a particular number of attributes. In adaptive strategies initial meanings start out quite holistic expressing many attributes. Generally it takes more time when meanings need to shrink to minimal size, which is the optimal lexicon for a non-structured world (see Figure 5.4c). In the non-structured condition the population has established a fully atomic lexical system which is indeed more difficult to arrive at given an adaptive strategy.

Structure thus helps the agents in bootstrapping their lexicon which captures the most frequently co-occurring attributes. However even in the structured condition over half of the meanings consist of only a single attribute. This can be understood by the way attributes are removed and added. Attribute removal is much more frequent because it happens whenever an attribute was expressed which was not part of the topic. Attribute addition only happens when the language game was a failure, which as shown in Figure 5.4a stops quite soon. So as soon as agents have reached full communicative success, the words only erode and never extend.

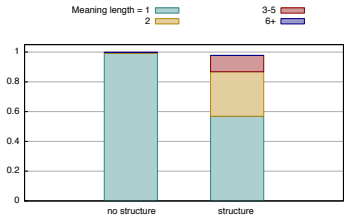
5.2.2 Impact of population size

Small increases in the population are not problematic with regard to communicative success as shown in Figure 5.5a. For a total of 100 attributes, 10 attributes per object, structure of 100 co-occurrence edges we see that scaling the population from 5 to 30 agents does not impact communicative success dramatically. The decrease in communicative success simply reflects the fact that a convention needs longer (i.e. more games) to be heard by all agents. For example in a population of five, an agent aligns six times as much as an agent in a population of 30 for the same amount of total games.

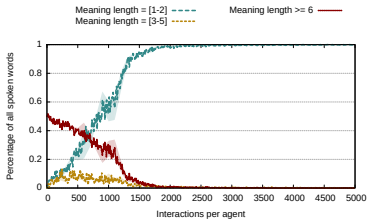
What is surprising is that although all populations eventually reach full or very high levels of communicative success they do not all reach the same high levels of lexical alignment as shown in Figure 5.5b. Alignment starts with a



(a) Communicative success

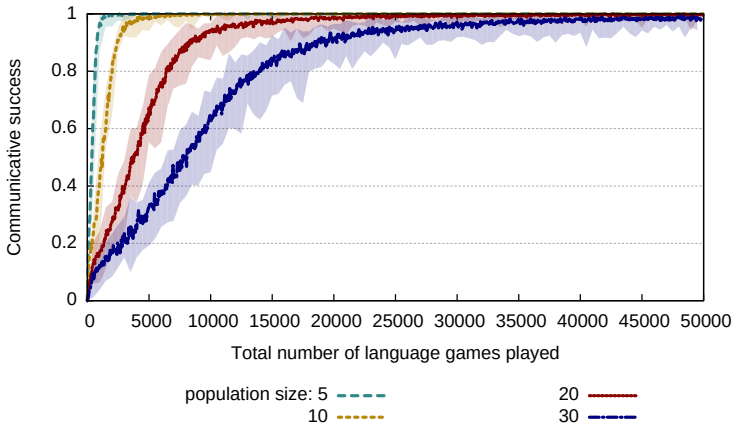


(b) Meaning length

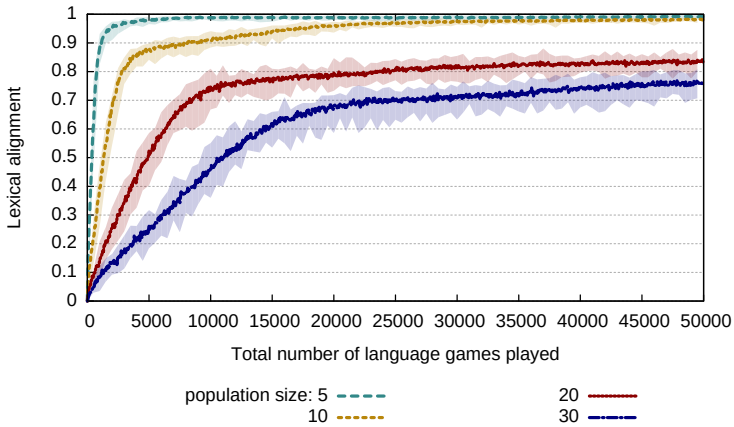


(c) Meaning length evolution

Figure 5.4: Impact of context structure on the Baseline Adaptive Strategy. (a) The Baseline Adaptive Strategy achieves full communicative success for both setups but structure helps in bootstrapping communicative success. (b) Each bar represents the percentage of words that express a certain amount of attributes. In the non-structured world the lexicon erodes to full atomicity. (c) Shows how the agents arrive at the fully atomic lexicon shown in the left-hand bar in (b),(a-c) Total number of attributes: 100, attributes per object: 10, edges: 0 and 100 population size: 10, number of simulations: (a) 20 (b-c) 4, error: 5 to 95 percentile, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$.



(a) Communicative Success



(b) Lexical Alignment

Figure 5.5: Impact of an increase in the size of the population. (a) and (b) Total number of attributes: 100, 100 co-occurrence edges, population sizes: 5, 10, 20 and 30, attributes per object: 10, co-occurrence edges: 100, number of simulations: 20, error: 5 to 95 percentile, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$.

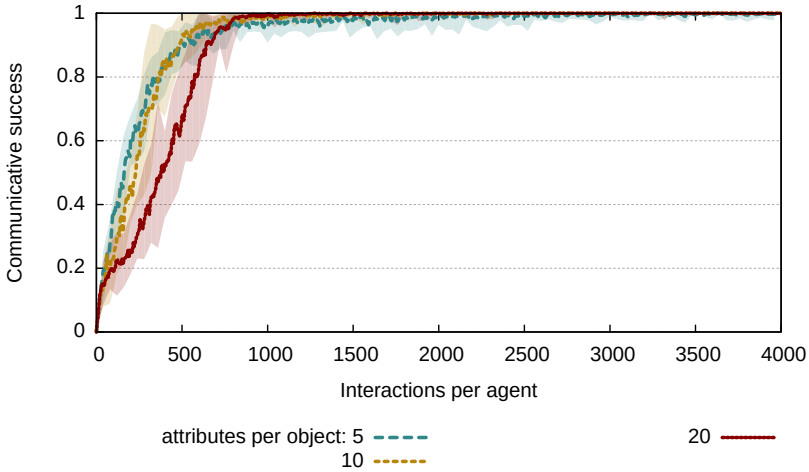


Figure 5.6: Communicative success for conditions with different attributes per object. Population size: 10, error: 5 to 95 percentile, number of simulations: 20, total number of attributes, 50, 100, and 200. $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$.

rapid increase followed by a much slower but still increasing alignment. Both stages can be interpreted as effects of fast and slow mapping (Carey & Bartlett, 1978). In the first phase of fast alignment agents are still adopting their initial meaning representation and fine-tuning the straightforward attributes. In the second phase the agents need to establish the more intricate details of the meaning which requires many exposures, especially with the rather crude shaping mechanics of the Baseline Adaptive Strategy. In the next section we introduce a strategy which shows more subtle and consistent alignment behaviour.

5.2.3 Impact of attributes per object

Increasing the number of attributes per object from 5 to 20 only impacts the communicative success marginally (see Figure 5.6). This is quite surprising since this parameter is the main regulator of the uncertainty the agents face. Assume a listener is exposed to a single new word and a topic of five attributes. The word could refer to any subset of the five attributes of the topic (except the empty set) which amounts to $2^5 - 1 = 31$, for twenty this becomes $2^{20} - 1 = 1048566$. This was one of the main reasons the Baseline

competitive strategy failed to scale to even four attributes per object. More attributes also entails that initial word meanings created in invention and adoption cover many more attributes and the meaning thus requires more shaping than a small one. Figure 5.6 however shows us the minimal effect this has on communicative success.

The reason why adaptive strategies cope rather well with the larger meaning spaces resulting from scaling up the attributes per object is twofold. First the similarity-based flexible processing allows a single meaning to capture a larger part of the meaning space. Second the adaptive shaping dynamics from alignment implement a more powerful search through the meaning space than competitive enumeration-based strategies do.

5.3 The Weighted Adaptive Strategy

Although the Baseline Adaptive Strategy embodies the core concepts of an adaptive strategy, namely flexible processing and adaptive representations, it does so in a minimal way. The baseline strategy shows potential with its ability to adapt to the structure of the world in interesting ways and can, in general, reach high levels of communicative success very fast. What the baseline strategy lacks is the ability to represent certainty or uncertainty about membership of an attribute to the meaning. In this section the baseline strategy is extended by adding the ability to maintain certainty weights for individual attributes.

5.3.1 Fine-grained adaptive meaning representation

The Weighted Adaptive Strategy extends the representational capacities of the agents by giving them the capacity to maintain a certainty score for each attribute of word meaning. Meaning is thus no longer a normal set, but a weighted set in which every element is scored. This representation is strongly related to both fuzzy set theory (Zadeh, 1965), with the degree of membership interpreted as the degree of (un)certainty, and prototype theory (Rosch, 1973). Because similarity plays such a crucial role, adaptive strategies are highly compatible with the ideas about conceptual spaces presented by Gärdenfors (2000, 2004). Because of small differences between fuzzy sets and the sets used here we call them *weighted sets* (Wellens *et al.*, 2008; Wellens & Loetzsch, 2012).

A weighted set M_w is represented as a normal set M and a certainty function $f_M(a)$ which associates a number in the interval $]0,1]$[2] to every

[2]The interval for fuzzy sets is $[0,1]$. Weighted sets do not include elements for which

element a in M. The certainty function $f_M(a)$ represents the certainty or uncertainty whether attribute a is part of meaning M. How $f_M(a)$ is calculated and updated is discussed later during alignment. Most common set operations are taken from fuzzy set theory.

union: $A_w \cup B_w = A \cup B$ with $f_{A \cup B}(x) = \text{Max}[f_{A_w}(x), f_{B_w}(x)],\ x \in A \cup B$

difference: $A_w \setminus B_w = A \setminus B$ with $f_{A \setminus B}(x) = f_{A_w}(x),\ x \in A \setminus B$

cardinality: $|A_w| = \sum_{x \in A} f_{A_w}(x)$

The intersection operation is changed from its fuzzy set variant, instead of taking the regular intersection and the minimum membership function it takes the regular intersection with the membership (certainty) function of the left operand:

intersection: $A_w \cap B_w = A \cap B$ with $f_{A \cap B}(x) = f_{A_w}(x),\ x \in A \cap B$.

Note that just like in the Baseline Adaptive Strategy intersection is no longer commutative. The reasoning to resort to a non-commutative intersection operation remains the same as for the Baseline Adaptive Strategy but for clarity is explained a bit more in the following section on processing.

With the ability to internally keep track of meaning (un)certainty and allowing agents to modify these scores based on their use, the representation becomes *adaptive* in a much more gradual fashion. Figure 5.7 shows two examples of this representation, the meaning of the artificial word "wabazu" and a toy-like meaning for the English word "dog". The attributes are not competing as they were in a cross-situational strategy for the Minimal Guessing Game (see chapter 3).

5.3.2 Flexible processing

As in the baseline strategy the similarity measure plays a pivotal role in processing. The similarity measure [equation 5.1] now uses the weighted set operations and cardinality and as such takes into account the certainty weights.

$$\text{Overlap}_w(M_w, O) = \frac{|M_w \cap O| \cdot |O \cap M_w| - |M_w \setminus O| \cdot |O \setminus M_w|}{|M_w| \cdot |O|} \tag{5.2}$$

their membership is 0. The reasons I chose to remove attributes with zero certainty is that otherwise words would become cluttered with attributes for which anyway their certainty is zero. These attributes anyway do not influence processing.

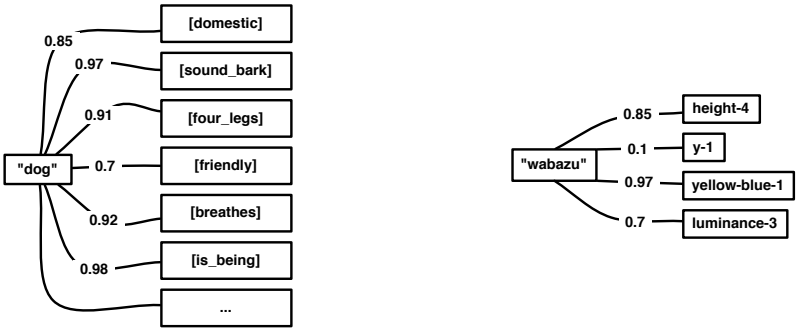


Figure 5.7: Representations for possible words in the Weighted Adaptive Strategy. On the left a fragment of how the meaning of "dog" could be represented. On the right an artificial word that was created in the experiments. Remark that attributes have a certainty score. For example [y-1] is associated to "wabazu" with only 0.1 denoting that the agent is quite uncertain whether it is part of the conventional meaning for "wabazu".

For objects O which are not weighted sets we assume that $f_O(x) = 1, \forall x \in O$. Intuitively we want the similarity measure to be large (close to 1) when the meaning M_w and the object O share many attributes, especially if they have high certainty scores. The shared attributes are calculated and weighed by $|M_w \cap O| \cdot |O \cap M_w|$ The larger the intersection between the meaning and the object and the higher the weights of these attributes in M_w the larger the result[3]. The weight of the disjoint attributes, calculated by $|M_w \setminus O| \cdot |O \setminus M_w|$, gets subtracted from the shared attributes. Only when the disjoint part is larger than the shared part can the similarity become negative. Because of the multiplication in $|M_w \setminus O| \cdot |O \setminus M_w|$, the similarity can never be negative when $|M_w \subseteq O|$. Some examples:

`Similarity(((a 1) (b .5) (c .7)), (a b c))` $= \frac{3.6 - 0}{3.6} = 1$
`Similarity(((a .5) (b .9) (c .3)), (d e f))` $= \frac{0 - 5.1}{5.1} = -1$
`Similarity(((a .9)), (a b c))` $= \frac{0.9 - 0}{2.7} = \frac{1}{3}$
`Similarity(((a .5) (b .5) (c .5)), (a c d))` $= \frac{2 - 0.5}{4.5} = \frac{1}{3}$
`Similarity(((a .9) (b .1) (c 1)), (a c d))` $= \frac{2 - 0.1}{3.3} = 0.58$

[3]Here it matters to use the non-commutative intersection as it will now allow the weights of the object (which in fact are all 1) to weigh in as they do when taking the set difference. If the commutative fuzzy intersection would be taken then the differences would weigh in more than the intersection as the fuzzy intersection would only count the minimum weight of the intersecting elements.

Finally we extend the similarity function so that given an utterance U and an object O, $\text{Overlap}_w(\text{U,O}) = \text{Overlap}_w(\bigcup_{w \in U} f_{meaning}(w), O)$ where the union $\bigcup$ takes the fuzzy union.

Production and interpretation are exactly the same as in the Baseline Adaptive Strategy except that the Overlap metric is replaced with the weighted version explained here. Production is a search process which, given a context C and a topic T, finds the utterance as follows:

$$\text{Produce(C,T)} = U \text{ that maximizes } \text{Overlap}_w\text{(U,T)}$$

Just like for the Baseline Adaptive Strategy, the flexibility allows the agents to use (combinations of) words that do not fully conform to the meaning to be expressed. Langacker (2000a)[p. 5] calls this cognitive capacity *extension*: "An act of categorization may also register some disparity between the categorizing structure and the target. In this case I speak of extension."

5.3.3 Invention and adoption

Invention is identical to that of the Baseline Adaptive Strategy (see Algorithm 14, page 148) except that each attribute in the initial meaning of the new word receives a low initial certainty score of 0.05, which is the default initial certainty score.

Like invention, adoption works much like that of the Baseline Adaptive Strategy with again the difference that the certainty function for each attribute of the initial meaning is initialized to 0.05 (see Figure 5.8). Starting out with such a low certainty score makes sense because the agents are indeed highly uncertain about the attributes of their first initial guess. These initial meanings are thus very fragile first approximations and generally quite specific in nature. Low certainty scores also allow the agent more flexibility in use.

5.3.4 Alignment

Alignment now takes into account the scoring capabilities of the agents. The first step of form competition remains the same as in the baseline strategy with two words seen as competitors when their associated meanings consist of the same attributes regardless of their scores.

Flexible word use entails that in a usage event some parts of the expressed meanings were beneficial (the shared ones) and others were not (the disjoint ones). Word use where meanings are stretched is called *extension* (Langacker, 2000a). An extended use entails some form of "strain" which in turn affects the meanings of these linguistic items. This is operationalised by slightly

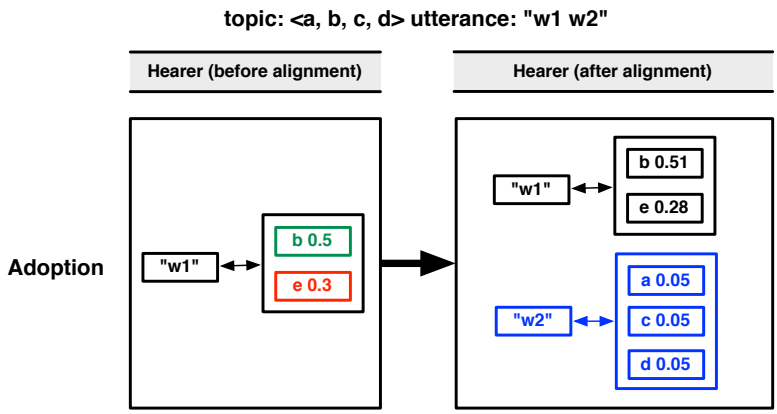


Figure 5.8: Schematic representation of adoption and the initial meaning. The topic consists of attributes ⟨a,b,c,d⟩ and the utterance "w1 w2" is spoken. The figure shows the adoption of word "w2" and its initial meaning.

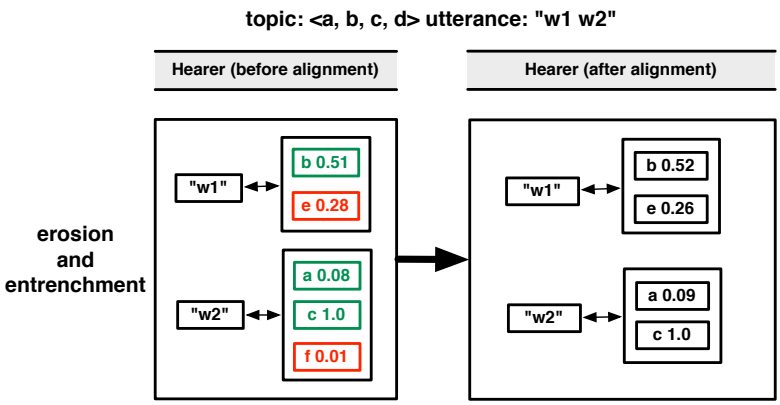


Figure 5.9: Schematic representation of erosion and entrenchment. The topic consists of attributes ⟨a,b,c,d⟩ and the utterance "w1 w2" is spoken.

shifting the certainty scores $f_M(a)$ every time a word is used in interpretation. The certainty score $f_M(a)$ of attribute a in meaning M is based on two statistics, the amount of matches *match* of that attribute with the topic attributes and the amount of mismatches *mismatch*. For example if *match* is 500 it means that attribute a in meaning M was also present in the topic when it was used for 500 times. If *mismatch* is 300 it means that 300 times attribute a was not part of the topic. Their sum (in this case 800) thus gives the total amount of times that attribute was associated to the word when it was used.

Both statistics are updated at the end of each language game and $f_M(a)$ is calculated as follows:

$$f_M(a) = \frac{\text{match} - \eta.\text{mismatch}}{\text{match} + \text{mismatch}} \tag{5.3}$$

The counters `match` and `mismatch` are initialized not to 0 or 1 but to 10 in order to dampen large fluctuations in the initial calculations[4]. The erosion parameter η determines the weight of mismatches and thus the speed of erosion. In all experiments $\eta = 1.5$. The more an attribute matches with the expressed topic the higher $f_M(a)$ which resembles the psychological phenomenon of *entrenchment*(Langacker, 2000a). Likewise, the more it mismatches the lower $f_M(a)$ until it reaches its lower limit 0 upon which the attribute is removed form the meaning. Decreasing scores resembles semantic erosion (also referred to as semantic bleaching or desemantisation). Entrenchment and erosion are shown in Figure 5.9.

Entrenchment and erosion cannot enlarge the initial meaning so the listener needs a way to add new attributes when necessary. Only in failed games when the listener knew all the spoken words does the listener decide to add attributes to the spoken words. Just like in the Baseline Adaptive Strategy the listener adds all unexpressed attributes of the topic, again with very low certainty scores, to all uttered words, further specifying those words. For more details see Algorithm 16 and Figure 5.10.

Combining similarity-based flexibility with entrenchment and semantic erosion, word meanings gradually shape themselves to better conform with future use. Repeated over thousands of language games, the word meanings progressively refine and shift, capturing frequently co-occurring features (clusters) in the world, thus implementing a search through the hypothesis space, and capturing only what is functionally relevant.

[4]This actually means my calculations in the example I gave earlier are not 100% correct. If *match* is 500 then it means it was only matched 490 times since the counter starts at 10.

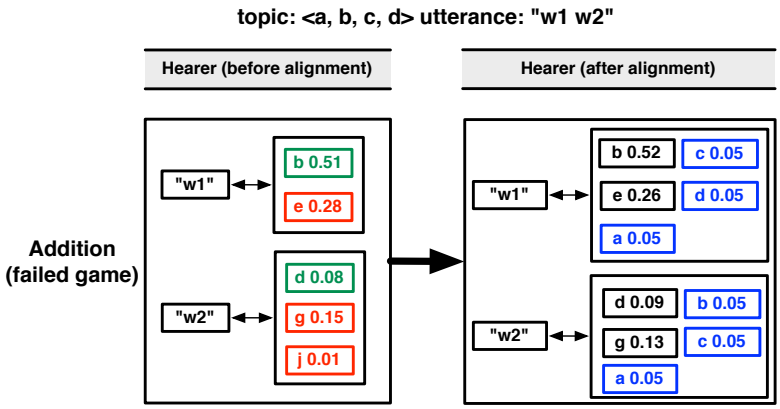


Figure 5.10: Schematic representation of addition of attributes. The topic consists of attributes ⟨a,b,c,d⟩ and the utterance "w1 w2" is spoken.

```
———————— Function Align_w(usedWords, topic, agent) ————————
// Form competition is the same as in baseline strategy

sharedAttributes ← f_meaning(utterance) ∩_w topic;
disjointAttributes ← f_meaning(utterance) \_w topic;

// Update certainty scores
ForEach word in usedWords
   ForEach attribute in f_meaning(word)
      If attribute ∈ sharedAttributes
      then IncrementScore(word, attribute);
      Else DecrementScore(word, attribute);
        remove attribute if score <= 0;
If not CommunicatedSuccessfully(agent)
then // Make words more specific, only the listener does this
   ForEach word in usedWords
      do Associate disjointAttributes to word; // low init. scores
———————————————————— End Align_w ————————————————————
```

Algorithm 16: Function Align_w(usedWords, topic, agent) updates the certainty scores of the used words. It increments the score for attributes that matched with the topic and decrements for those that didn't as explained at the beginning of this Section. When a certainty scores becomes negative its association is removed. In case of a failed game the listener also adds new attributes to the used words.

5.4 Experimental results

In order to measure lexical alignment for the Weighted Strategy we need to adapt the alignment measure used in the Competitive Strategies and the Baseline Adaptive Strategy [see page 115, equation (4.1)]. Reusing the metric $\texttt{Overlap}_w$ is not advised because it was designed specifically for comparing meanings with objects (of weighted sets with normal sets) whereas for lexical alignment we compare meanings with meanings. The metric used is the following:

$$\text{Alignment}(A,B) = \frac{|A \cap_w B| + |B \cap_w A|}{|A| + |B|} \tag{5.4}$$

The addition of weights positively impacts the convergence speed of communicative success and especially that of alignment (see Figure 5.11a). The subtle and slower updating mechanisms at the level of an individual agent speed up the alignment at the level of the population. Not only does it speed up alignment, the agents also reach a higher level of alignment. Although the connection between alignment and communicative success is not trivial, better alignment should prove more robust in case the communicative encounters become more demanding.

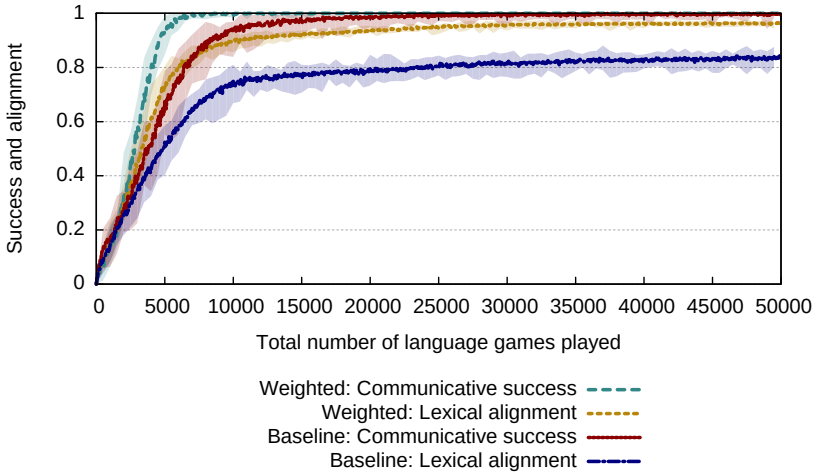
One of the weaknesses of the Baseline Adaptive Strategy was the low alignment with an increasing population (see Figure 5.5b, page 156). In Figure 5.11b we see that this is no longer the case using the weighted strategy. The addition of weights thus makes the approximation of each others word meanings much more accurate and does not suffer drastically from an increase in population size.

5.5 In-depth example of a language game

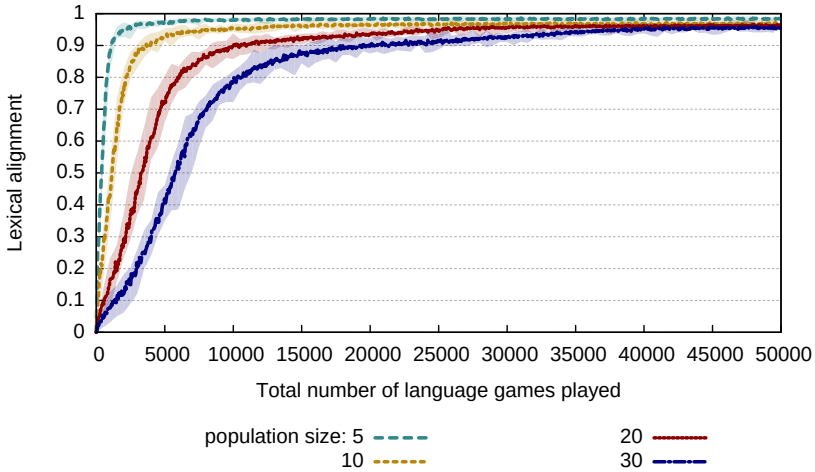
In this section I want to give a more in-depth example of one language game in which the agents implement the Weighted Adaptive Strategy. For this example I use a new type of data source from which I generate objects to communicate about.

5.5.1 Communicating about mushrooms

So far we have only tested adaptive strategies on data sets that have been generated by the algorithm explained in Chapter 4, Section 4.5.2. In the next chapter we investigate the difficulties and requirements of using robotic data. Before turning to robotic data I want to give an in-depth example of one language game in which the agents implement the Weighted Adaptive Strategy. The data used to generate objects is no longer the earlier used



(a) Communicative success and lexical alignment



(b) Impact of population size on lexical alignment

Figure 5.11: (a) Communicative success and lexical alignment for a population of 20 agents following the Baseline and the Weighted Adaptive Strategy. (b) Evolution of lexical alignment when increasing the size of the population. (a) and (b) Total number of attributes: 100, population size: (a) 20 (b) 5-30, error: 5 to 95 percentile, number of simulations: 20, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$.

algorithm but one of the larger UCI Machine Learning data sets (Frank & Asuncion, 2010; Wellens, 2008). The data set, called the mushroom data set, contains 8124 mushrooms, each described by 22 multi-valued features of which we only kept the following 12 (including the class as an attribute)[5]:

cap shape: bell, convex, flat, knobbed, sunken

cap color: brown, buff, cinnamon, gray, green, pink, purple, red, white, yellow

odor: almond, anise, creosote, fishy, foul, none, pungent, spicy

gill color: black, brown, chocolate, gray, green, pink, red, white

stalk shape: enlarging, tapering

stalk root: bulbous, club, cup, equal, rhizomorphs, rooted, missing

stalk surface above ring: fibrous, scaly, silky, smooth

stalk surface below ring: fibrous, scaly, silky, smooth

stalk color above ring: brown, gray, pink, red, white

spore print color: brown, buff, chocolate, green, purple, white

population: abundant, clustered, numerous, scattered, several, solitary

class: edible, poisonous

The Compositional Guessing Game requires Boolean predicate-like attributes so the multi-valued features are converted into predicates. For example a mushroom that has value flat for the attribute cap-shape, is converted to the Boolean attribute "cap-shape=flat". This brings the total number of Boolean attributes to 58 with 12 attributes per object.

The data set was originally developed for classification tasks (edible versus poisonous) (Iba *et al.*, 1988). Here we are not interested in classification but draw instances from the data set to generate contexts for agents to communicate about. In the following experiments a population of 25 agents played language games with objects being taken from the mushroom data set. In each game the participating agents are confronted with a context of 5 to 15 mushrooms, of which one is the topic. All results are averaged over 4 runs with error-bars showing minimum and maximum values.

[5]This data set is also artificially generated based on a much smaller set of actual mushrooms.

5.5.2 Example language game

As an example of the Weighted Adaptive Strategy without ordering constraints I now present an example taken from an experimental run in which a population of five agents is in the process of establishing a shared lexicon as they have only played a total of 1000 language games. The speaker of the game currently has a lexicon of 45 words and needs to express a topic with the attributes as shown in Figure 5.12

Topic		
class=e	population=v	spore-print-color=k
stalk-color-above-ring=w	stalk-surface-below-ring=s	stalk-shape=t
gill-color=u	odor=n	cap-color=e
cap-shape=x		

Figure 5.12: The attributes of the topic of the game.

word form	attributes (meaning)	certainty score (expr/unexpr)	similarity
	stalk-surface-below-ring=s	0.66 (46/1)	
	odor=n	0.68 (43/0)	
"sezasa"	stalk-shape=t	0.64 (42/1)	0.59
	class=e	0.68 (43/0)	
	cap-shape=x	0.48 (34/4)	
	spore-print-color=n	0.05 (22/13)	
	stalk-color-above-ring=w	0.08 (12/6)	

Figure 5.13: The first words "sezasa" and its attributes with their certainty score. In brackets are the counters on which the score is based. Similarity of the meaning with the topic is 0.59.

The speaker starts by proposing the first word "sezasa" (see Figure 5.13). This representation alone uncovers quite some details about the history of this lexical entry. The bracketed numbers after the certainty score express how many times the attribute was part of the topic (matched) when the word was heard and the other the amount of mismatches. Their sum represents the total amount of times the agent has heard this word with this attribute associated as part of its meaning. From this we learn that at some point only the attribute [stalk-surface-below-ring=s] was associated and since then the word has been heard 47 times with this attribute associated to it, 46 times it

word form	attributes (meaning)	certainty score (expr/unexpr)	combined similarity
"dumomi"	odor=n	0.33 (18/3)	0.75
	stalk-shape=t	0.39 (19/2)	
	stalk-surface-below-ring=s	0.33 (18/3)	
	spore-print-color=k	0.09 (14/7)	
	class=e	0.45 (20/1)	
	cap-shape=f	0.04 (12/7)	
	population=v	0.20 (7/1)	
	stalk-color-above-ring=g	0.10 (2/0)	
"fawaso"	cap-color=e	0.47 (25/2)	0.84
	odor=n	0.52 (26/1)	
	stalk-shape=t	0.52 (26/1)	
	stalk-surface-below-ring=s	0.52 (26/1)	
	class=e	0.57 (27/0)	
	spore-print-color=n	0.12 (12/5)	
"nixiso"	gill-color=u	0.32 (15/2)	0.934
	spore-print-color=n	0.17 (6/1)	
"raroso"	spore-print-color=k	0.36 (23/4)	0.937
	stalk-surface-below-ring=s	0.25 (16/4)	
"giwolo"	odor=n	0.40 (23/3)	0.940
	stalk-shape=t	0.35 (22/4)	
	stalk-surface-below-ring=s	0.35 (22/4)	
	class=e	0.40 (23/3)	
	stalk-color-above-ring=w	0.31 (17/3)	
	spore-print-color=n	0.11 (6/2)	

Figure 5.14: Remaining words in the utterance in the memory of the speaker. Blue attributes added positively to the similarity because they matched and were not yet expressed by a previous word. Green attributes contribute as well because they increased the certainty of an already expressed attribute. Red attributes mismatched and decrease similarity.

word form	attributes (meaning)	certainty score (expr/unexpr)
"sezasa"	stalk-surface-below-ring=s	0.54 (28/1)
	odor=n	0.54 (37/3)
	stalk-shape=t	0.62 (39/1)
	cap-shape=x	0.48 (34/4)
	class=e	0.67 (40/0)
"dumomi"	stalk-surface-below-ring=s	0.40 (20/2)
"fawaso"	cap-color=e	0.39 (22/3)
	population=y	0.17 (18/7)
	odor=n	0.37 (12/0)
	stalk-surface-below-ring=s	0.55 (24/0)
	spore-print-color=k	0.10 (4/1)
"nixiso"	gill-color=u	0.39 (16/1)
	stalk-color-above-ring=g	0.17 (6/1)
"raroso"	spore-print-color=k	0.61 (31/0)
	stalk-surface-below-ring=s	0.32 (12/1)
	population=v	0.07 (10/5)
	stalk-color-above-ring=w	0.09 (9/4)
"giwolo"	odor=n	0.29 (18/4)
	stalk-shape=t	0.40 (20/2)
	stalk-surface-below-ring=s	0.40 (20/2)
	class=e	0.40 (20/2)
	stalk-color-above-ring=w	0.36 (17/2)
	cap-shape=f	0.36 (14/1)
	cap-color=n	0.14 (11/4)
	cap-color=g	0.10 (2/0)

Figure 5.15: Representations for the words in the utterance by the listener. Attributes in red are not part of the speaker's meaning for the same word. This does not mean they are "wrong", it might be that they become part of the conventionalized meaning. Also note that for the word "dumomi" the listener associates only one attribute whereas the speaker associates eight.

was also part of the topic and only once it was not. Four exposures later the agent must have been part of an unsuccessful game with this word because he added three attributes [odor=n], [stalk-shape=t], [class=e]. In the same way he added the three remaining attributes. Only one attribute mismatches with the topic (spore-print-color=n) resulting in a similarity of 0.59 between topic and the word meaning of "sezasa". Note that the mismatching attribute has the lowest certainty of all the associated attributes. The speaker then continues to add another 5 words resulting in the utterance "sezasa dumomi fawaso nixiso raroso giwolo". Figure 5.14 shows in detail how each word increased combined similarity.

In this particular game the listener knows each of the spoken words, but as the population is still in the process of negotiating the meanings he associates not entirely the same attributes to each of those words as shown in Figure 5.15. Although the meanings of speaker and listener have not yet fully aligned, the listener is able to point to the correct object and the game is a success.

5.5.3 The impact of mutual exclusion and attribute addition

Using the same UCI data set we first investigate the impact of mutual exclusion. In invention and adoption, the agents rely on the constraint of mutual exclusion when constructing an initial meaning by excluding the attributes already expressed by other words in the utterance. By removing mutual exclusion the agents employ a holistic approach when constructing the initial meaning. By keeping all experimental parameters identical we can identify the impact of mutual exclusion on communicative success and the nature of the emergent lexicon.

The second parameter under investigation is the addition of attributes by the listener in failed games. With addition removed the agents can only change their meanings through erosion and meanings can thus only become more general over time. We are interested in this parameter because in general the addition of semantic content receives much less attention than the loss of semantic context (i.e. erosion). For example Langacker (2000a) explains in detail the phenomena of erosion and entrenchment but not what could trigger the addition of new semantic content. For this reason we want to see how much the addition of attributes influences the emergent lexicon.

Figure 5.16 shows that communicative success remains almost identical with or without any of the two parameters active. Either these parameters do not influence the emergent lexicons or the high levels of communicative success are due to the flexible nature of processing which enables the agents to cope with different types of lexicons.

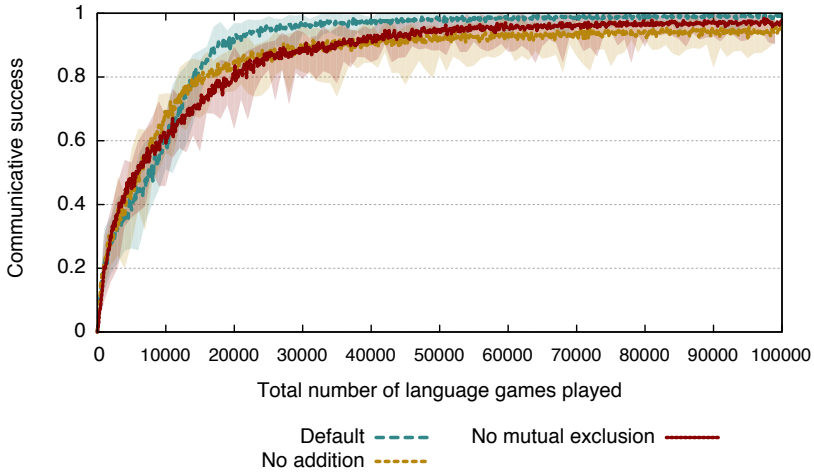
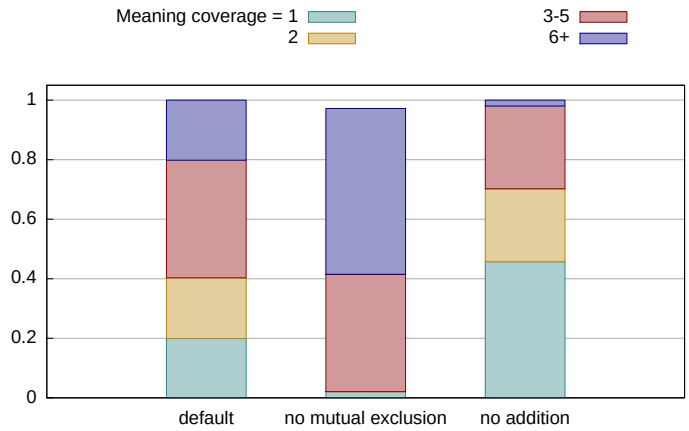


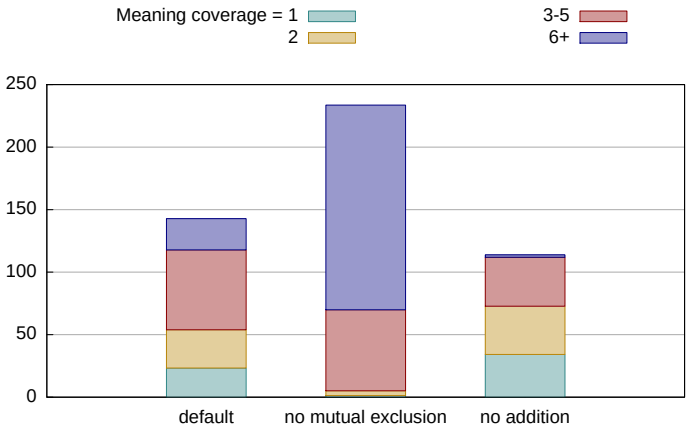
Figure 5.16: Impact of mutual exclusion and the addition of attributes on communicative success. population size: 25, context size: 5-15 mushrooms, error: min/max, number of simulations: 4, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$.

That the meanings expressed by the emergent lexicons do indeed differ is shown in Figure 5.17. Figure 5.17a illustrates that without the constraint of mutual exclusion atomic word meanings or even meanings expressing two attributes are almost non existent. Specific meanings have become far more dominant. The initial meaning thus influences the final conventionalized meaning considerably. Not only that but Figure 5.17b shows that without mutual exclusion the emergent lexicon is also much larger. This makes sense, when the agents have more specific words they also need more of them. Note that they still manage to require less than 250 words to communicate successfully about over 8000 mushrooms.

We see the opposite results when it is not possible to make meanings more specific (add attributes). Words with atomic meanings take up almost half of all spoken words. Specific words expressing over 6 attributes are almost non-existent and used only very infrequently. What is interesting is that not *all* word meanings are eroded to become atomic in nature. The words expressing more than one attribute could maintain that configuration although no addition of an attribute was ever allowed. This means all the attributes is still expresses must have been part of the initial meaning.



(a) Percentage of attributes expressed per meaning



(b) Lexicon size subdivided by expressiveness

Figure 5.17: Impact of mutual exclusion and the addition of attributes on the nature of the emergent lexicon.(a) Each stacked bar shows the percentage of words that express 1 (atomic), 2, 3-5, or over 6 attributes. (b) The stacked bars show the total size of the emergent lexicon (averaged over the population) again divided by the amount of attributes expressed. (a) and (b) population size:25, error: min-max, number of simulations: 4, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$.

5.6 Conclusion

In this chapter I presented a cross-situational strategy that takes an adaptive instead of a competitive approach to the problem of meaning uncertainty. The competitive strategies have primarily been developed to solve the problem of mapping uncertainty and not the more difficult problem of meaning uncertainty. When confronted with multi-dimensional meanings and compositional utterances, agents have to establish the internals of the meanings themselves instead of only their mappings to the correct word form.

Competitive and adaptive strategies can be seen as supplying a computational model for two different views on word learning. Competitive strategies rather subscribe to the view in which the bulk of the meaning (or concept) formation has already been done and all that remains to be done is mapping words to these pre-established meanings. Adaptive strategies support a view in which a child or learner is seen as gradually shaping and constructing the meanings based on new exposures in new contexts.

As such adaptive strategies are still cross-situational in that they need to contrast their meaning hypothesis with new exposures, new objects, new contexts. Rather than using these situations to eliminate competitors, they use it to internally shape and adapt their prototypical meaning representations.

Adaptive strategies start from the same observation made by the Discriminative Competitive Strategy (Section 4.7). Both acknowledge that the key problem lies in the enumeration of competing hypotheses. The Discriminative Strategy takes the approach to try and minimize this enumeration but further keep true to the tenets of competitive (enumeration-based) learning. Adaptive approaches take a bolder approach and do away with the enumeration altogether.

Both the Baseline and the Weighted Adaptive Strategy maintain only a single prototype-like hypothesis. This is only the first of three crucial steps which constitute the definitional features of an adaptive strategy. The second requirement is that adaptive strategies have to support flexible processing. In the Compositional Guessing Game, where meanings are represented as sets, this is achieved by an overlap-based similarity measure. For more complex representations, such as feature structures, more advanced types of flexibility are required. The crucial feature of the flexibility is that it should allow a linguistic item to be used even though it not fully conforms to the intended conceptualization or meaning. Agents thus need to be able to compare two representations and based on what is shared and what is not shared make a judgement of compatibility or similarity.

The role and importance of such a flexible comparison in learning has been studied by many people. Children acquire verb meaning more easily when

there is possibility of comparison (Childers & Tomasello, 2001; Childers, 2008; Childers & Paik, 2009). Comparison helps adults and children to acquire relational concepts and language (Gentner *et al.*, 2003; Loewenstein & Gentner, 2001; Wang & Baillargeon, 2008). Gentner and Namy (Gentner & Namy, 1999, 2004; Namy & Gentner, 2002; Gentner & Medina, 1998) have shown that comparison processing allows children to notice commonalities between objects that are not noticed when the objects are examined in isolation.

The third and final definitional feature is that in alignment the feedback from flexible processing should be taken into account. In the Baseline Adaptive Strategy this was achieved by removing mismatching attributes and under certain conditions adding new attributes to the word meanings of the spoken words. In the Weighted Adaptive Strategy this process became a bit more subtle by being able to increment or decrement individual certainty weights per attribute.

The Weighted Adaptive Strategy shows remarkable overlap with the usage-based dynamic model that Langacker (2000a) proposes. For example, increments of certainty weights can be seen as *entrenchment* which Langacker describes as "The occurrence of psychological events leaves some kind of trace that facilitates their re-occurrence." Later on Langacker also introduces the phenomenon of *abstraction* which he describes as "the emergence of a structure through reinforcement of the commonality inherent in multiple experiences. By its very nature, this abstractive process "filters out" those facets of the individual experiences which do not recur." Although this description could also apply to competitive cross-situational learning it is clear from the context that Langacker is referring to a process that is abstracting the internals of meanings and not the removal of competitors. This type of emergent usage-based clusters are observed in the lexical systems resulting from adaptive shaping dynamics.

Another process fundamental to cognition, according to Langacker (2000a), is "the ability to compare two structures and detect any discrepancy between them." He goes on to say "... an act of categorization may also register some disparity between the categorizing structure and the target. In this case I speak of extension". The flexible processing in adaptive strategies is thus very compatible with the notions Langacker sees as fundamental. Later he even captures the essence of the adaptive shaping dynamics as follows: "Lexical items arise through a process of progressive decontextualization, where non-recurring aspects of usage events are filtered out through lack of reinforcement."

I thus see adaptive strategies, at least at the lexical level, to stay true to the most important aspects of the cognitive usage-based model proposed by Langacker (2000a). See Table 5.1 for an overview of the strategies for the

	Baseline Competitive	Discriminative Competitive
(1) full CS	low uncertainty	yes
(2) high alignment	low uncertainty	yes
(3) usage-based	low uncertainty	no
(4) schematicity	low uncertainty	no
(5) scoring	yes	yes
(6) solves WFC	no	no
	Baseline Adaptive	Weighted Adaptive
(1) full CS	yes	yes
(2) high alignment	small populations	yes
(3) usage-based	yes	yes
(4) schematicity	yes	yes
(5) scoring	no	yes
(6) solves WFC	no	no

Table 5.1: Overview of Compositional Guessing Game strategies. Following questions are addressed: (1) Can the strategy reach full communicative success? (2) Does it reach high levels of lexical alignment? (3) Do the emergent lexicons show usage-based characteristics? (4) Do the emergent lexicons show schematicity (general and specific word meanings) (5) Does the Strategy implement a scoring mechanism (not counting word form competition scoring). (6) Does the strategy by itself also solve the problem of word form competition? If no, it means it requires the addition of a Naming Game strategy. The slots in the table are either yes or no, or state the condition under which it can be reached, a sort of "yes, but only when ...".

Compositional Guessing Game so far.

In the evaluation and comparison of the strategies for the Compositional Guessing Game I have tried to measure not only for communicative success and lexical alignment but also for specific characteristics in the emergent lexicons. The most important of these measures was the amount of attributes expressed per word meaning.

The Baseline Strategy could reach communicative success very rapidly even when the population size was increased. Lexical alignment on the other hand did not scale that convincingly with an increasing population size. High lexical alignment is not a necessity, definitely not when full communicative success can be reached anyway. However it does mean that the shaping dynamics of the Baseline Adaptive Strategy do not lead to high or full levels of alignment when full communicative success has been reached. The Weighted Adaptive Strategy did show a further lexical alignment even after reaching full communicative success even when population size was increased.

I also showed that it was more difficult for a population to reach full communicative success when the world showed no structure (re-occurring patterns). They did reach it, but only after eroding all the word meanings to an atomic meaning. When no attributes tend to co-occur among the object then an atomic language is most optimal, even from a usage-based perspective. Although as soon as structure is introduced we saw, even with the Baseline Adaptive strategy, the emergence of more varied lexicons which then also helped in bootstrapping the lexical system.

The emergent lexicons thus showed both specific and general (atomic) words. The more atomic word meanings take longer to be established and in general start out as more specific words. This type of change in meaning is also observed in natural languages. For example the French color word "orange" comes from the fruit "orange". The word orange thus underwent a parallel process of gradual erosion so that only its color attributes remained. It became only popular to denote the color in the mid sixteenth century. Obviously in these simulations these transitions occur much faster, but nevertheless the same type of erosion mechanisms might lie at its root.

In the next chapter we further investigate the weighted adaptive strategy using robotic data. There we further investigate the dynamics of gradually shaping word meanings and shed more light on the issues regarding schematicity and the effects of extension and erosion.

Chapter 6

Robustness in Compositional Guessing Games

In the previous two chapters we have explored competitive and adaptive strategies to the problem of meaning uncertainty as posed by the Compositional Guessing Game. All experiments so far have either used simulated worlds in which we control how contexts are structured or machine learning data sets. In this chapter we investigate how both types of strategies cope when robotic data instead of simulated data is used. In particular the problem of having different views on the same context poses a new and interesting problem to the compositional guessing game and provides a perfect opportunity to test the robustness of competitive and adaptive strategies (Loetzsch & Spranger, 2010).

6.1 Interacting autonomous robots

The robotic setup used in this experiment is similar to other experiments that investigate the cultural transmission of language in embodied agents [e.g. Spranger, 2012, Steels & Loetzsch, 2008]. The experimental setup involves at least two QRIO robots (Fujita *et al.*, 2003) with the ability to perceive physical objects in a shared environment using their cameras, to track these objects persistently over time and space and to extract attributes from these objects. The robots must establish joint attention (Tomasello, 1995) in the sense that they share the same environment, locate some objects in their immediate context, and know their mutual position and direction of view. Finally, there have to be non-linguistic behaviors for signaling whether a communicative interaction was successful and, in case of failure, the robots need to be able to point to the topic object.

The robots maintain continuous and persistent models about the sur-

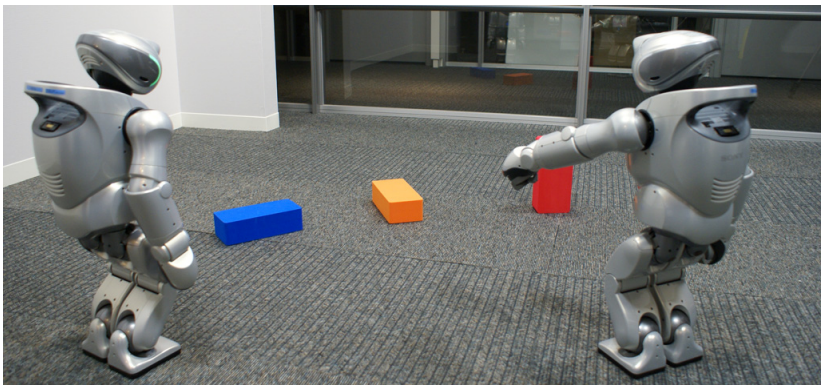
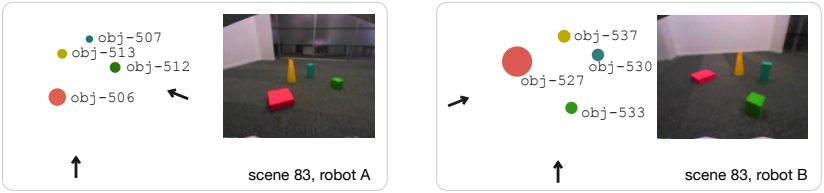


Figure 6.1: Sony QRIO humanoid robots play a language game about physical objects in a shared scene.

rounding objects using probabilistic modeling techniques (Röfer *et al.*, 2004; Spranger, 2008). As a result, each agent has a representation of every object in the scene, including estimated position, size and color properties (see the top of Figure 6.2). From each such model, values on seven continuous *sensory channels* are extracted, being the position of the object in an egocentric coordinate system (`x` and `y`), the estimated size (`width` and `height`), the average brightness (`luminance` or `lum`), average color values on a green/red and a yellow/blue dimension (`grn-red` and `yell-bl`). Optionally also the uniformity of the brightness and color values within the object (as the standard deviation of all pixels within the object region in the camera image) and the maximum and minimum values of these color channels can be added, resulting in a total of sixteen continuous channels per object. Channel values are scaled between the interval of 0 and 1, which is then split into four regions, a technique that could be compared to discrimination trees (Steels, 1998a; Smith, 2001, 2005a). One out of four Boolean attributes is assigned to an object for each channel according to the intervals of each channel value. For example the green/red value for `obj-506` in Figure 6.2 is `0.88`, so the assigned attribute is `grn-red-4`. We refer to the list of objects with their associated attributes as *sensory context*.

Just like in the previously reported language games all agents start with empty lexicons and have never before seen any of the physical objects in their environment. At the beginning of a game the agents establish a *joint attentional scene* (Tomasello, 1995) – a situation in which both robots attend to the same set of objects in the environment and register the position and



obj-512	obj-507	obj-513	obj-506	obj-533	obj-530	obj-537	obj-527
width-2	*width-1*	width-2	*width-3*	width-2	*width-2*	width-2	*width-4*
height-2	*height-2*	height-4	height-2	*height-3*	*height-3*	height-4	height-2
lum-2	*lum-2*	lum-3	lum-2	lum-2	*lum-3*	lum-3	lum-2
grn-red-2	grn-red-1	grn-red-3	grn-red-4	*grn-red-1*	grn-red-1	grn-red-3	grn-red-4
yell-bl-2	yell-bl-3	yell-bl-1	*yell-bl-2*	yell-bl-2	yell-bl-3	yell-bl-1	*yell-bl-3*
x-3	x-4	x-4	*x-2*	*x-2*	x-4	x-4	*x-3*
y-1	*y-2*	*y-3*	*y-3*	*y-2*	*y-1*	*y-2*	*y-4*

Figure 6.2: Visual perception of an example scene for robot A and B. On the top, the scene as seen through the cameras of the two robots and the object models constructed by the vision system are shown. The colored circles denote objects, the width of the circles represents the width of the objects and the position in the graph shows the position of the objects relative to the robot. Black arrows denote the position and orientation of the two robots. On the bottom, a subset of the attributes that were extracted for each object are shown. Since both robots view the scene from different positions and lighting conditions, their perceptions of the scenes, and consequently the attributes extracted from their object models, differ. Attributes that are different between the two robots are printed in italics.

orientation of the other robot. Once such a state is reached, the game starts. One of the agents is randomly assigned to take the role of speaker and the other the role of listener. Both agents perceive a sensory context (as described above) from the joint attentional scene. The speaker randomly picks one object from his context to be the *topic* of this interaction. His communicative goal is to describe the topic such that the listener is able to point to it. He thus constructs an utterance, inventing new words when necessary. These mechanisms are described in detail in the previous two chapters (Section 4.7, page 127 for the Discriminative Strategy and Section 5.3.2, page 159 for the Weighted Adaptive Strategy). The listener parses the utterance using his own lexicon and points to the object from his own perception of the scene that he believes to be most probable given the utterance. In case the listener did not point to the correct topic, the speaker will point to the object he intended. Otherwise he signals success (by nodding his head). At the end of the game the listener knows the intended topic, but not the subset of attributes the

speaker chose to express and certainly not how the words themselves relate to the subsets. Finally, at the end of each interaction both agents modify their lexicons slightly based on the sensory context, the topic and the utterance (see Section 4.7.2, page 129 for alignment of the Discriminative Strategy and Section 5.3.4, page 161 for the Weighted Adaptive Strategy.).

Since conducting thousands of such language games with real robots would be very time-consuming and also because we wanted repeatable and controlled experiments, we recorded the perceptions of the two robots (as in Figure 6.2) for 150 different scenes, each containing between two and four different objects of varying position and orientation out of a set of ten physical objects. A random scene from this collection is then chosen in every language game and the two different perceptions of robots A and B are presented to the two interacting agents. In these simulations, agents point to objects by transmitting the x and y coordinates of the objects (in their own egocentric reference system). The agent receiving these coordinates can transform them into a location relative to its own position using the offset and orientation of the other robot.

Of particular interest is the impact of the different perceptions both agents have on the same scene. The agents have no apriori knowledge about which sensory channels are more or less prone to be categorized differently by the other agent. For example, the x and y coordinates have a tendency to differ frequently since they are relative to the robot's own position. The three color channels on the other hand tend to be shared more often. A robust strategy should thus maintain high levels of communicative success even when perceptions partly differ and preferably the emergent lexicons should reflect the stable tendencies in the perceptions.

In what follows populations with both the Weighted Adaptive Strategy and the Discriminative CS Strategy are tested on the robotic data. We start with the adaptive strategy where we pay considerable attention to how the adaptive nature helps in shaping and aligning the emergent lexicon. We then compare these results with the competitive strategy under different conditions. We test for the influence of the different perceptions, of the number of sensory channels and of the size of the population.

6.2 Experimental results for the Weighted Adaptive Strategy

We tested the Weighted Adaptive Strategy introduced in the previous chapter with a population of 25 agents that play repeated series of 50000 language games (see Figure 6.3). From early on (at around interaction 10000), the

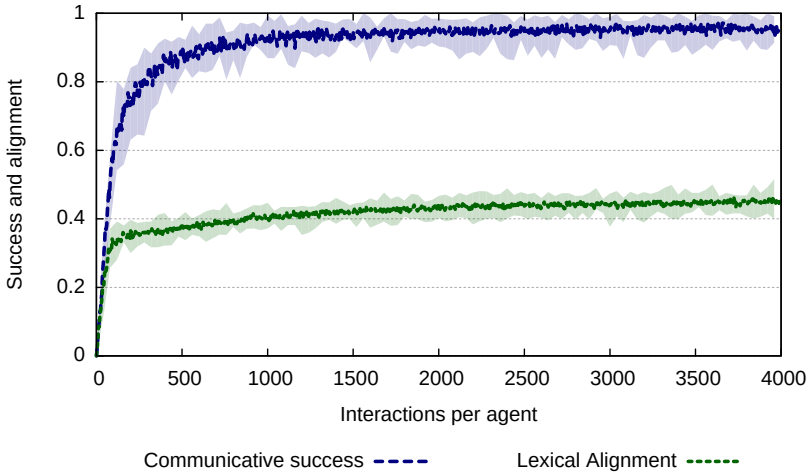


Figure 6.3: Dynamics of the language games in a population of 25 agents averaged over 12 runs of 50000 interactions. Values are plotted for each interaction along the x-axis. Communicative success: For each successful interaction (the listener understands the utterance and is able to point to the object that was chosen as topic by the speaker), the value 1 is recorded, for each failure, 0. Values are averaged over the last 100 interactions. Lexical alignment was introduced in the previous chapter (page 165). A value of 1 means that all 25 agents have identical lexicons, -1 means that they are completely different (each agent associates completely different attribute sets to each word form). Error bars represent max and min values across the 12 different experimental runs.

agents communicate successfully in approximately 90% of the cases. Note that on average each of the 25 agents takes part in only 800 out of 10000 interactions and thus play only 4000 games in total. Although the agents communicate successfully almost from the start, lexical alignment does not follow the same trend. Low lexical alignment indicates that the agents continue to associate non-identical attribute sets to their words.

The reason the agents do not reach full communicative success is due to the different perceptions of the agents as shown by Figure 6.4. We ran the exact same experiment again but now "cheated" by providing both agents with the visual perception of only one robot. In this scenario the agents do reach full communicative success and also reach a much higher level of alignment. This shows that the different perceptions increase the challenge

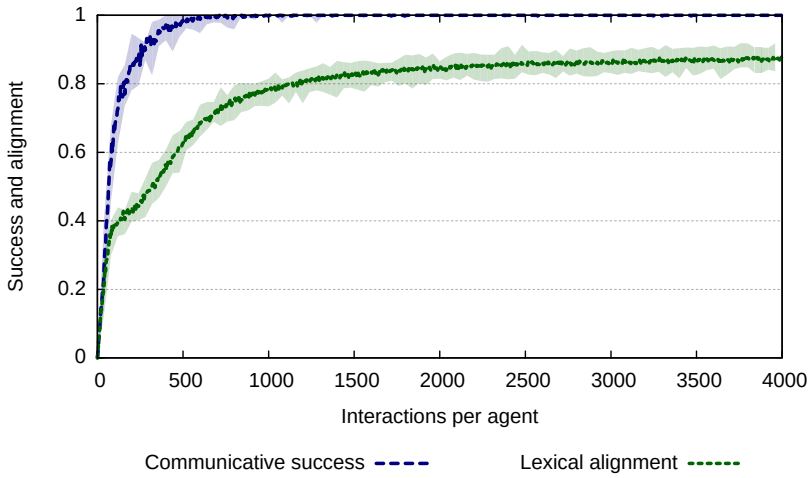


Figure 6.4: Impact of different perceptions on communicative success and lexical alignment. All experimental parameters are identical to those of the graph shown in Figure 6.3 except that both agents share the same perceptions of the scene.

of bootstrapping and aligning a lexicon considerably and that the Weighted Adaptive Strategy is susceptible to these difficulties, yet manages to still reach acceptable levels of success.

6.2.1 Understanding alignment

In order to better understand the progression of alignment of the adaptive strategy, Figure 6.5 lists the meanings of the first three words of agent 1 after 10000 interactions (communicative success $\approx$ 90%) and compares them with the meanings that agents 2 and 3 connect to these forms. For each word, the associated attributes to it and the scores of the association are shown (sorted by score). It is immediately clear why lexical alignment remains rather low in the population: each agent indeed associates drastically different attribute sets of highly varying size to the same word forms. For example, all three agents associate different height information to the word "murifo": none for agent 1, `height-4` and `height-3` for agent 2 and `height-3` for agent 3. The number of attributes connected to the word "nuside" ranges from three (agent 3) up to eight (agent 1). For nearly every word form, each agent associates at least one attribute that no other agent connects to the same form. Words

form	agent 1		agent 2		agent 3	
"murifo"	x-4	0.46	lum-2	0.57	lum-2	0.38
	lum-2	0.40	yell-blue-4	0.40	x-4	0.29
	std-lum-2	0.25	grn-red-2	0.40	height-3	0.25
	y-3	0.19	x-4	0.32	std-grn-red-2	0.13
	std-yell-blue-2	0.03	height-4	0.19	std-yell-blue-3	0.08
			std-grn-red-2	0.12	yell-blue-4	0.08
			std-yell-blue-2	0.10	y-3	0.08
			std-grn-red-1	0.10		
			std-lum-2	0.10		
			height-3	0.10		
			width-4	0.10		
			y-3	0.10		
"nuside"	lum-2	0.58	yell-blue-2	0.68	yell-blue-2	0.56
	yell-blue-2	0.49	lum-2	0.59	std-yell-blue-1	0.27
	std-yell-blue-1	0.39	width-2	0.31	height-3	0.24
	std-lum-1	0.24	std-yell-blue-1	0.29		
	height-3	0.19	std-lum-1	0.17		
	std-grn-red-1	0.17	x-3	0.17		
	y-3	0.17	std-grn-red-2	0.08		
	x-4	0.17				
"migata"	grn-red-2	0.50	lum-2	0.40	lum-2	0.44
	lum-2	0.48	std-lum-2	0.33	x-4	0.38
	yell-blue-4	0.39	std-grn-red-2	0.32	std-lum-2	0.21
	std-lum-2	0.33	std-yell-blue-3	0.32	yell-blue-4	0.20
	std-yell-blue-3	0.30	grn-red-2	0.32	grn-red-2	0.10
	std-grn-red-2	0.22	x-4	0.32		

Figure 6.5: Evolution of alignment. The meanings of the first three words of agent 1 ("murife", "nuside" and "migata") (out of a population of 25 agents) and the corresponding meanings in the lexicons of agents 2 and 3 after 10000 interactions. The numbers on the right side are scores of the association to the attribute.

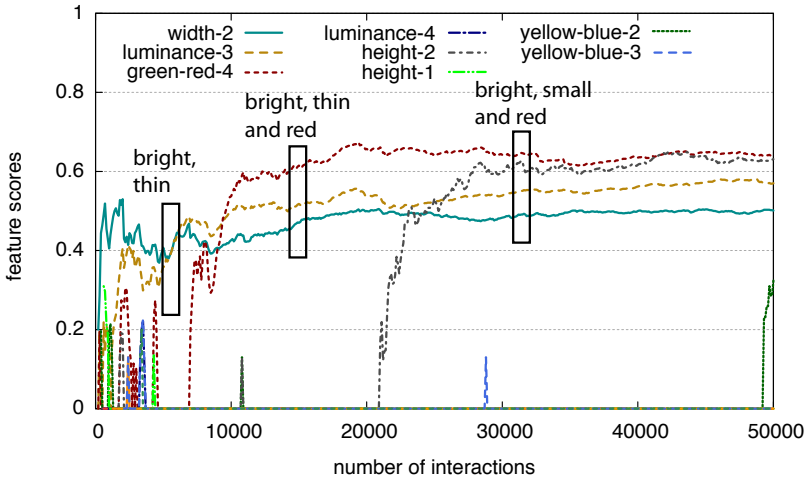
can even be associated to multiple attributes on the same sensory channel. For example, agent 2 has the attributes `std-grn-red-1` and `std-grn-red-2` in its attribute set for the word "murifo".

The agents could not, however, communicate successfully if word meanings were not (at least) partially shared. Despite all the differences, the meanings of the three words in Figure 6.5 start to emerge: (almost) all agents associate `x-4`, `y-3` and `lum-2` to the word "murifo", giving it the meaning "far, left, uniformly dark". For "nuside", the attributes `yell-blue-2`, and `std-yell-blue-1` are shared by all three agents and taking only the first two we can add `lum-2`, `std-lum-1` to that, resulting in a shared meaning "rather dark and uniformly yellow" for agents 1 and 2 and just "uniformly yellow" for agent 3. The third word "migata" is associated by all three agents with `grn-red-2`, `lum-2` and `std-lum-2` which means "dark green". This level of alignment is already enough for the agents to communicate successfully in many different contexts and shows that the meaning of the words is not just the set of associated attributes but lies more in how the words are *used*. Alignment continuously increases during the remaining 40000 interactions (see Figure 6.3), allowing the agents to communicate successfully in 95% of the cases after 50000 interactions.

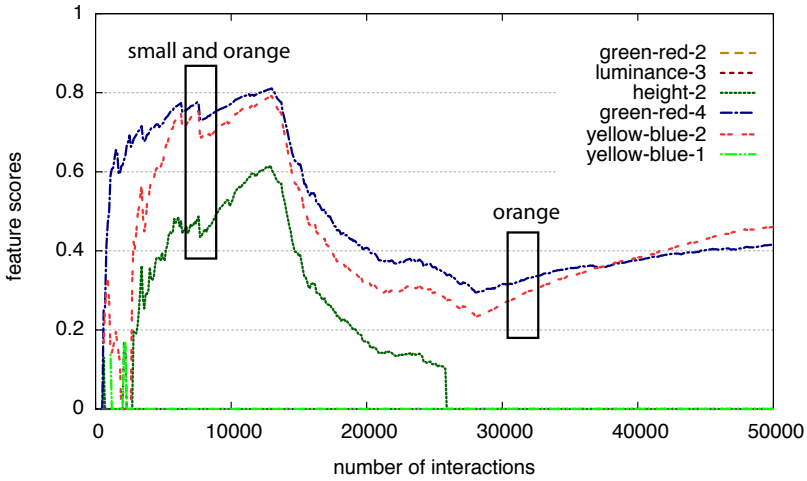
In order to understand how the agents are able to align their initially differing lexicons, we looked at how the meaning of a single word in one agent evolves over time. The alignment dynamics continuously adapt the certainty scores of the associated attributes and as such shape the meaning as a whole over the course of many usage events. Figure 6.6 gives two examples of the shaping of word meanings over time. Despite some other associations that disappear very quickly, the word in Figure 6.6a is initially only connected to `width-2`. Over the course of many interactions, more attributes are associated (`luminance-3` at around interaction 3000, `green-red-4` at interaction 7000 and finally `height-2` at interaction 22000). The meaning of this word thus changed from general ("thin") to very specific ("thin, low, bright and red"). The word in Figure 6.6b is an example of the opposite. It starts out very specific, with connections to `green-red-4`, `yellow-blue-2`, `height-2`, ("orange and small") and loses most of these attributes, becoming more general ("orange") towards the end.

6.2.2 Fast mapping

As pointed out in earlier chapters human learners can infer usable, yet fragile, meanings for a novel word after only a few exposures, a phenomenon called *fast mapping*. The previous results do not give us any insight on this issue, as all results so far depicted a population in the process of bootstrapping



(a) General to specific



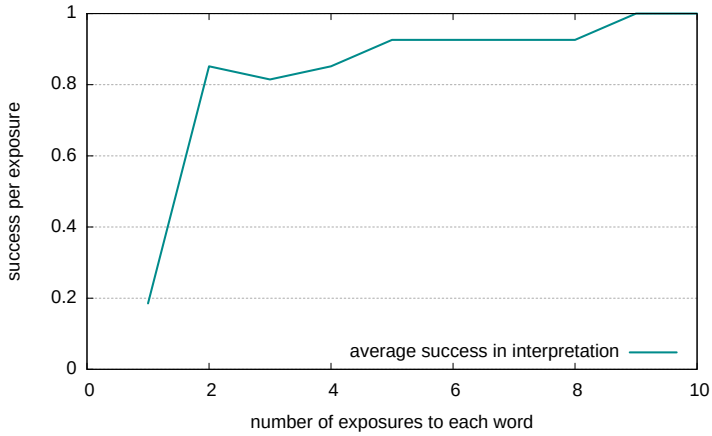
(b) Specific to general

Figure 6.6: Adaptive word meanings. Each graph shows, for one particular word in the lexicon of one agent, the certainty score of the associated attributes. In order to keep the graphs readable, the agents agents only have access to a subset of 5 sensory channels (`width`, `height`, `luminance`, `green-red`, `yellow-blue`).

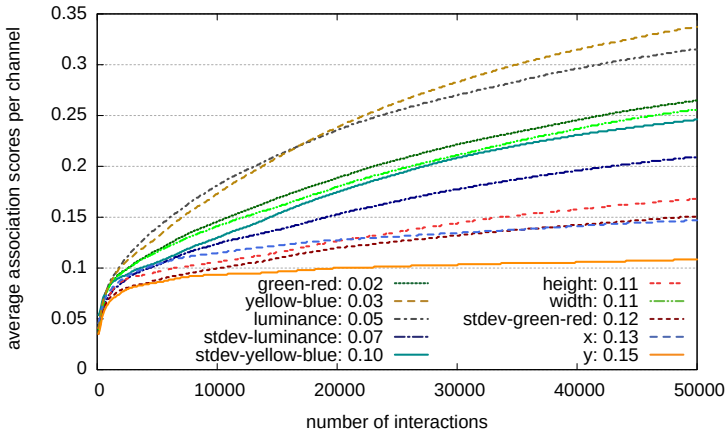
a lexicon. To investigate whether an adaptive agent's initial meanings also show high usability early on, we added a new agent to a population that had already conventionalised a shared lexicon. The new agent only takes the role of a listener, resembling a child in a population that speaks a fairly stable language. The results, as depicted in Figure 6.7a, show that by the time of the second exposure 85% of the novel words lead to a successful interpretation. Further exposures gradually improve this result and by the tenth exposure all words result in a successful interpretation. This is even more surprising given that the other members of the population are unaware they are talking to a new agent, and thus use multi-word utterances, making it harder for the new agent to grasp the meanings of the words. In 20% of the cases, the new agent successfully interprets the utterance on the very first exposure to a new word because he understands enough of the other words to be able to point correctly.

6.2.3 Impact of different perceptions

When agents are embodied in physical robots, they have to deal with perceptual noise. The two robots view the scene from different angles and under different lighting conditions, leading to different perceptions of the same physical objects. However, the similarity in perception varies depending on the sensory channel. The average distance between the perception of a physical object between robots A and B on each sensory channel is shown in the legend of Figure 6.7b. This distance is computed by iterating over all objects of all scenes in the data set and for each sensory channel averaging the distances of the sensory values between the two robots. From the result we see that the most reliable sensory channels are `green-red` (average distance 0.02), `yellow-blue` (0.03) and `luminance` (0.05). The most varied channels show a very high level of difference, which makes them less suitable for successful communication: `y` (0.15), `x` (0.13) and `std-green-red` (0.12). Figure 6.7b shows, for each sensory channel, the strength with which attributes are associated. This average score is computed for each channel by iterating over all the words in the population and averaging the scores of connections to attributes on that channel. The highest average scores are for attributes on the `yellow-blue`, `luminance` and `green-red` channels, the lowest for attributes on `y`, `x` and `std-green-red`. This corresponds very well to the average sensory differences on these channels, showing that the agents cope with perceptual differences by relying less on unreliable channels.



(a)



(b)

Figure 6.7: a) The interpretation performance of one new agent that is added to a stabilised population. For each word this agent adopts, the communicative success at the first, second, third etc. exposure is measured and averaged over all the words in the lexicon of that agent. b) The impact of the different perceptions on the lexicon: for each sensory channel, the average association score for channel attributes is shown, given all words in the population. In the legend, for each channel the average difference between the perception of robots A and B for all scenes in the data set are shown.

6.2.4 Comparison with the Discriminative CS Strategy

From the competitive strategies in chapter 4 we can only test the discriminative variant on the embodied data. The robotic data contains from 5 to 16 attributes per object to which the baseline competitive strategy does not scale. In order to compare with the Adaptive Strategy we run the discriminative strategy with the same parameters as those for Figure 6.3. We see in Figure 6.8a that for a population of 25 agents and 10 attributes per object the agents cannot reach acceptable levels of communicative success.

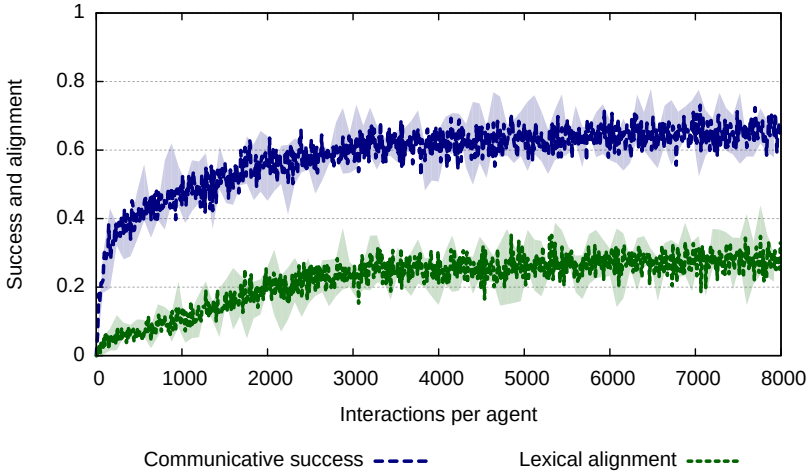
This low performance cannot be due only to the size of the population or the number of attributes because in Chapter 4 (e.g. Figure 4.11 on page 131) we have seen that the strategy can reach success with the same parameter settings in a simulated world. One hypothesis is that the problems lie in the different perceptions the agents have on the scene as discussed in Section 6.1. As before, we can test this by artificially letting the agents share their perception of the scene. Instead of each robot having his own visual perception, speaker and listener share the same view. Results in Figure 6.8b show that indeed when the robots share the same view, the population can establish a high rate of communicative success.

Although the difference in perception is the main culprit as to why the discriminative strategy cannot reach communicative success the discriminative strategy still requires a large amount of interactions to reach success in the condition with shared perception as compared to the Adaptive Strategy (see Figure 6.9).

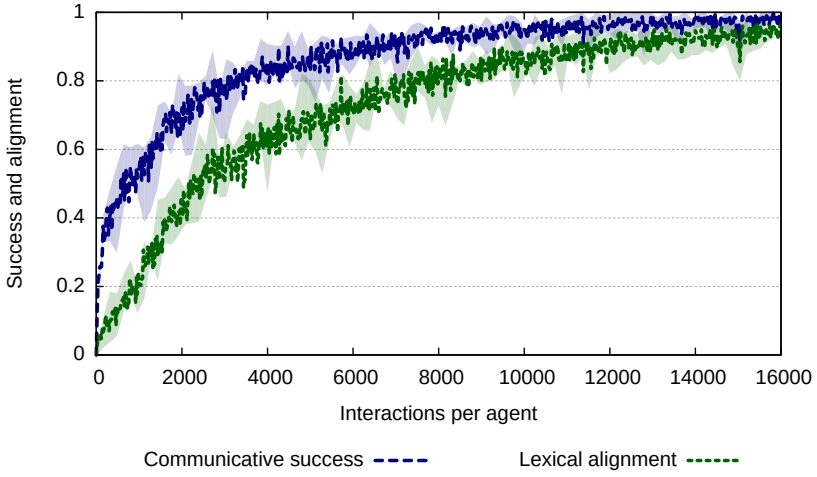
In order to ascertain whether it is the number of attributes per object or the size of the population which makes discriminative strategy slower, we ran another experiment where agents share the same perception but varied the size of the population (Figure 6.10a) and the number of sensory channels per object (Figure 6.10b). The Weighted Adaptive Strategy outperforms the Discriminative strategy along both dimensions. This is quite surprising because the Discriminative Strategy reduces the problem to a mapping problem while the Adaptive Strategy does not.

6.3 Conclusion

This chapter did not introduce a new type of strategy nor did it introduce an entirely new conventionalization problem. It did introduce a more difficult version of the meaning uncertainty problem, namely a version with non-equal alignment sets. What this means is that there is a difference in the sets that speaker and listener are producing, interpreting and aligning for. In this



(a) Different perception



(b) Shared perception

Figure 6.8: Communicative success and lexical alignment for a population of 25 agents, 10 channels (total 40 attributes, 10 per object). (a) Each agent has his own visual perception of the scene. (b) The robots share their perception eliminating possible mismatches between detected attributes. (a+b) nr of simulations: 4, error: min/max

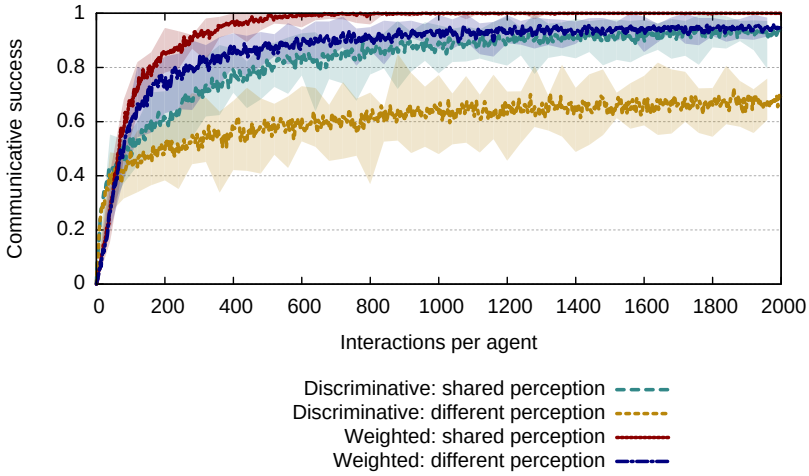
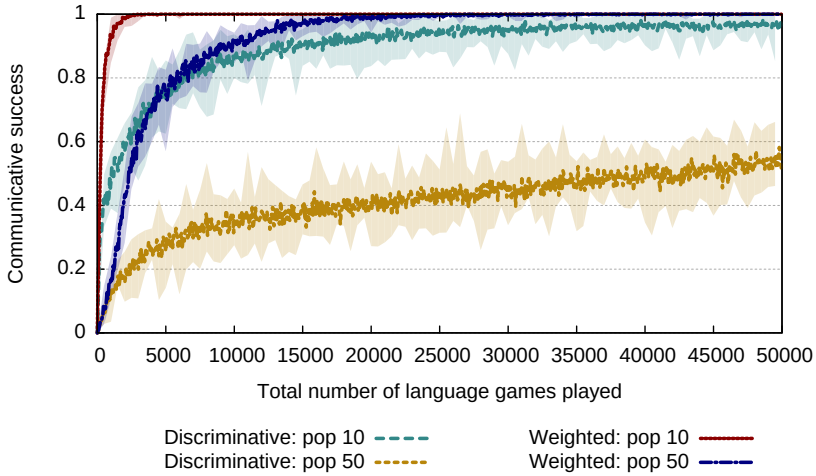


Figure 6.9: Communicative success for both the Discriminative and the Adaptive Strategy with and without shared perception. population size: 10, number of channels: 10 (40 total attributes, 10 per object). Nr of simulations: 12, error: min/max

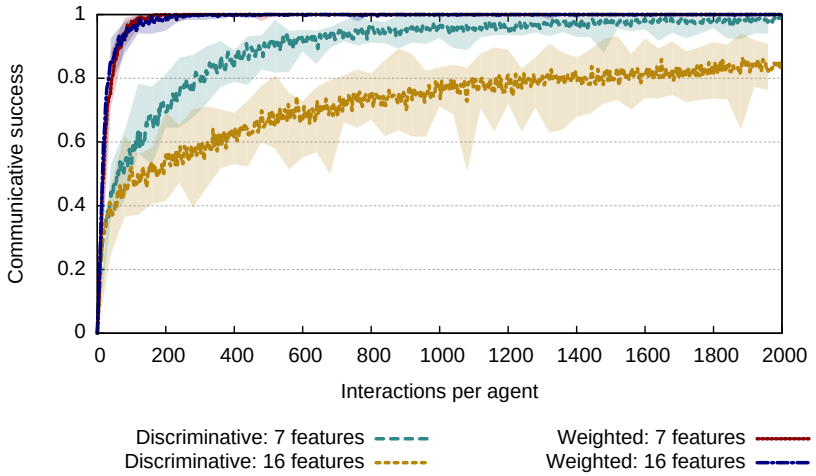
chapter I introduced this problem by using robotic data as the robots do not share the same view of the scene which indeed leads to different sets of attributes for the same objects. This problem could also have been introduced by other means than the two perspectives, for example by adding noise to the attributes lists.

One consequence is that the re-entrance test of the speaker is much less reliable since the speaker uses his own perception and does not know which sensory channels might be different for the listener. There has been work on implementing perspective reversal (Schober, 1993) to alleviate this difference in perspective (Steels & Loetzsch, 2008; Loetzsch *et al.*, 2008a) or one could introduce a more powerful a priori categorisation than cutting the channels in four regions. But in this chapter I was specifically interested in how competitive and adaptive strategies would deal with the problem of having not differences in their attribute sets.

The first results showed that these differences in view influence both communicative success and lexical alignment for the Weighted Adaptive Strategy. This could be verified by playing the same games but giving the robots a single, shared view. In this case the agents did reach full communicative success and high levels of alignment.



(a) Increase in population



(b) Increase in number of channels/attributes

Figure 6.10: Influence of population size and number of attributes. Along every tested dimension adaptive agents win over competitive agents. Especially the robustness in the number of sensory channels is remarkable. Nr of simulations: 12, error: min/max

Upon a further investigation of the alignment dynamics we showed that the different views do indeed lead to quite different attribute sets being associated to the same words by different agents. We also observed that among the differences also a core that was shared started to emerge. This core allowed them to reach communicative success (90%) whereas the differences explain the low alignment.

The reason the agents were able to reach high levels of communicative success even though they were confronted with different views is twofold. First the flexibility in processing is quite important when attribute sets are not the same among agents, especially in interpretation. But interestingly this was not the only reason. Upon a deeper investigation into the emergent lexicons we discovered that attributes that tend to be more reliable (like colors) were captured (entrenched) more than attributes that were unreliable (like position). This emergent effect made the lexical meanings themselves more robust to the different views.

We showed that word meanings can change from specific to general and the other way around. This is not surprising since adaptive strategies are compatible with some of the core principles of grammaticalization in which lexical items gradually shape into grammatical items (Hopper, 1991; Traugott & Heine, 1991; Heine *et al.*, 1991). The theory of grammaticalization shows with an abundance of evidence that grammatical meanings are gradually shaped through processes of extension, desemanticization, decategorialization and erosion (Heine & Kuteva, 2007). For example (Heine & Kuteva, 2007, p. 39) writes:

> Desemanticization is an immediate consequence of extension: Use of a linguistic expression E in a new context C entails that E loses part of its meaning that is incompatible with C – in other words, the two are Janusian sides of one and the same process.

This quote describes with remarkable accuracy the core principles of the adaptive strategies. Instead of desemanticization I have used the term erosion and like Heine, but in a more limited sense, used the term extension to denote a flexible type of processing where certain parts of the meaning mismatch with the intended meaning. In grammaticalization, however, the processes of extension and desemanticization are more complicated. For example extension is often triggered by metaphorical language use, which goes beyond the similarity based processing of the adaptive strategies. This is not to say that similarity has no role to play in human language processing. For example, Gentner & Rattermann (1991) stress the fundamental role similarity plays in natural languages and in analogical reasoning.

I also compared the Weighted Adaptive Strategy to the Discriminative Competitive Strategy and found that the Competitive Strategy could not

deal with the different attribute sets. Agents reached only about 60% communicative success and alignment remained very low (around 30%). This was no longer the case when agents could share the same view on the scene, confirming that the different views are the source of the failure.

Given that the agents could share the same perspective I also investigated the speed of convergence when scaling population size and the number of attributes (or dimensions). The Weighted Adaptive Strategy reached full communicative success for a population of 50 agents even before the Discriminative Competitive Strategy could reach full communicative success for a population of 10 agents. The same observation was made with respect to the number of attributes. The Weighted Adaptive Strategy felt almost no impact whereas the impact on the Discriminative Strategy was more profound. Again the adaptive strategy outperformed the discriminative strategy.

Part IV

Semantic categorization and conclusion

Introduction

In the previous two parts we have focused on purely lexical categorization. Agents create, adopt and shape lexical items, associations of a meaning with a word form. In the Naming Game and the Minimal Guessing Game meanings are taken to be holistic or are assumed to have been established earlier. In the compositional guessing game meanings exhibit internal structure and utterances are composed of multiple words. Natural languages, however, consist of multiple layers of categorization, beyond the lexical domain. Grammatical constructions, for example, further categorize words into semantic and syntactic subcategories.

In the first chapter of this part the topic shifts from lexical meaning to grammatical categories. The second and final chapter wraps up the main results and findings of the thesis and suggests avenues for future research.

In Chapter 7 I investigate the hypothesis whether adaptive strategies could be employed outside of the lexical domain. In the previous chapters we have used adaptive strategies to categorize the objects which the agents need to communicate about. Here we are interested whether the same mechanisms allow a population to align a second level of (semantically motivated) categories on top of the lexical categorization?

Adjectival ordering Sproat & Shih (1988) in natural languages is used as a source of inspiration for the experimental setup. In many natural languages, the order of adjectives is not random. For example "the indifferent Belgian consumers" sounds just fine, yet *"the Belgian indifferent consumers" sounds weird in English. It turns out that the ordering is based on the semantics of the adjectives. For example, in English, adjectives denoting a nationality always come closest to the noun.

The key problem in learning such an ordering lies in a feature that at first sight might seem rather harmless. The category-based slots are optional when producing a phrase. For example instead of "the indifferent Belgian consumers" I can also say "the indifferent consumers", not communicating the nationality. From a learner's perspective who is still figuring out the categories themselves this proves to be quite a challenge because every time a category is skipped he remains uncertain whether it was skipped or whether his own (fragile) category is incorrect or incomplete.

This question boils down to two specific research questions, which will become clearer in the chapter itself:

1. Can a population of agents employ an adaptive strategy to categorize the lexical meanings themselves? This means that they need align categories based on word meanings which are private and might not be perfectly aligned.

2. Since category-slots are optional can the resulting uncertainty be solved by agents employing adaptive strategies?

The answer to question (1) has already been partially answered in Chapter 6 where I showed that the Weighted Adaptive Strategy can align meanings when attribute sets are not shared. From a more abstract point of view agents face the same problem when creating internal semantic categories. The second question is new and leads to a new conventionalization problem, the problem of optionality.

By no means is the intention of that chapter to offer a fully fledged model of adjectival ordering. In fact even the syntactic categories of adjective and noun are missing from the lexicons so far. The emphasis lies on bootstrapping a second level of categorization and to shed light on the problems and uncertainty that come with it.

Chapter 7

Adaptive strategies in the domain of grammar

Construction grammar holds that grammatical constructions are not fundamentally different from lexical constructions in that they are also pairings of form and meaning (Fillmore *et al.*, 1988; Goldberg, 1995; Croft, 2003). The difference is that the expressed meaning tends to be relational in nature and is not expressed by words but by acting on words, for example by ordering them, or by marking, or adding stress or a different intonation pattern. Although grammatical constructions act on words, the constructions do not refer directly to specific words but rather to categories to which the words belong. These categories can be semantically motivated (e.g. animacy) or more syntactic (e.g. noun) in nature.

For example in the English utterance "Lucy wrote me a poem" one cannot just change the order of the words without changing or removing its meaning. In the active voice "Lucy" has to be in front of the verb "wrote" in order to be considered the actor of the writing. The exact same grammatical construction is involved in the utterance "Elfie baked me a delicious cake" which makes it clear that the underlying grammatical constructions abstract away from the actual lexical constructions like "wrote" or "bake" or "Lucy" or "Elfie". Grammatical constructions thus relies on categories to denote which set of lexical items (or constituents) can fill a particular slot.

7.1 Case study: Semantically motivated ordering constraints

A good example of semantic categorization is adjectival ordering in English or French. Adjectives are not added randomly in front or after the noun

but follow an ordering based on their semantics. For example Sproat & Shih (1988) propose the following simplified ordering classification for English:

$$Det > A_{quant} > A_{qual} > A_{size} > A_{form} > A_{color} > A_{nationality} > Noun \quad (7.1)$$

resulting in the following example orderings:

- many strange sharp Egyptian signs
- different fluffy little grey rabbits

In the experiments so far (e.g. in the Compositional Guessing Game) the order of the words has been largely random or determined by the production process of the strategy. For example in the similarity-based production algorithm of the adaptive strategies (Section 5.2) words that are most similar to the topic are found and consequently uttered first. This is very different from the type of ordering that we see for adjectives in two ways. First the ordering in the Compositional Guessing Game is *not* based on convention but just a consequence of the algorithm and second the ordering is not based on semantic categories. In each game the topic changes so the most similar words also change and there is no reason to assume that there is an underlying shared category among these words. In fact, the contrary seems more plausible.

The main questions discussed among linguists investigating adjectival ordering is whether this ordering shows universal tendencies or not (Sproat & Shih, 1991). For example is there a universal tendency to always put words denoting a nationality closest to the noun? With the current experimental setup I cannot contribute to this question. Instead my focus lies on the problem of how the semantic categories, categorizing the lexicon, can become established and aligned among a population of agents. I have shown that adaptive strategies can bootstrap and align rich multi-attribute word meanings, which categorize the objects in the world. This problem of bootstrapping and aligning a category-based word order pattern poses two new conventionalization challenges on top of those found in the lexical experiments.

To introduce these two problems I first want to refresh how an adaptive strategy gradually shapes the meanings of words. Let's look at alignment by the listener in a communicatively successful game in which the listener also knew all the spoken words (no adoption). During alignment the listener knows the topic and needs to align his set of used words. He achieves this by looking at the overlap (commonalities) and the differences between the topic and each single word, entrenching the commonalities and eroding the differences. This core process leads, over the course of thousands of games, to a gradual shaping of word meanings that become more and more aligned

in the population. It is fair to say that meanings are shaped "towards" the objects the agents need to communicate about. And because all agents shape, over the course of many games, towards the same objects, the meanings become aligned.

Calling it "shaping towards objects in the world" might be a bit too simplistic because the agents are actually shaping towards their experience (their set of attributes) of the topic object. Indeed, the goal of the previous chapter was to investigate whether shaping dynamics, and thus adaptive strategies, could cope with differing conceptualizations (sets of attributes) of the same topic. In short, can agents still reach communicative success and acceptable levels of alignment when they are "shaping towards non-equal attribute sets". The results showed that indeed they could, although not perfectly. Humans are known to have several cognitive capabilities that help them to get a more accurate estimate of the conceptualization of the speaker. Perspective reversal (Flavell, 1992) is one of them, Tomasello (2001) even speaks of intention reading.

When bootstrapping semantic categories that categorize the lexicon, agents face the same problem as they have with different conceptualizations. These semantic categories are no longer shaped "towards objects in the world" but "towards word meanings that fill the slot of the category". This will be discussed in more detail later but I hope this depiction is nevertheless clear. Adaptive strategies shape internal semantic categories by trying to make the categories more like the words that fit them. Every time a word fills in a particular category-based slot it has an impact (leaves a trace) on that category. The internal semantic category responds by slightly entrenching the commonalities and eroding the differences. But since word meanings are in many cases not 100% aligned, agents are shaping the categories towards differing attribute sets. This is the problem investigated in the previous chapter.

But there is a second problem, which has not yet been addressed and which follows from the problem of establisging a word order pattern. More specifically, the problem stems from the optionality of a category-based slot in processing. Taken ordering 7.1 from before, it is not required, and actually completely exceptional, that we would express a word for each of these categories each time we utter adjectives. For example the phrase "the intelligent Korean kid" expresses only $Det > A_{qual} > A_{nationality} > Noun$ and leaves $A_{quant} > A_{size} > A_{form} > A_{color}$ unexpressed.

When learning the categories underlying this ordering the optionality poses a significant difficulty. When a speaker decides to skip a category, he cannot signal this to the listener. When trying to map the utterance to his own category-based pattern the listener is confronted with the following

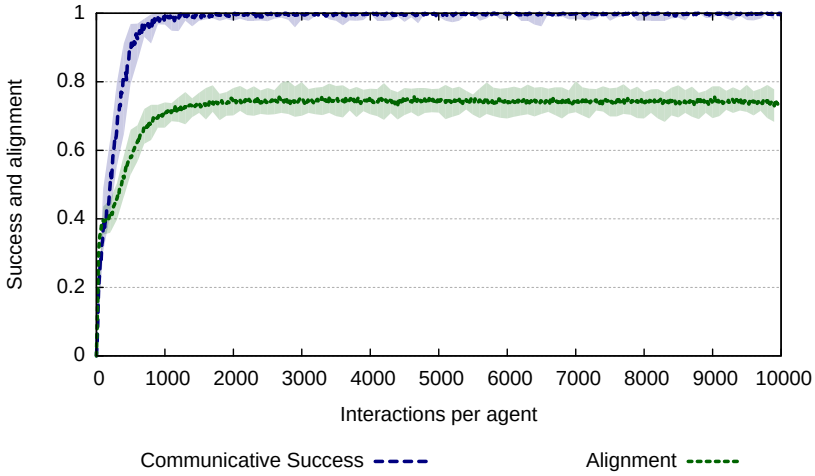


Figure 7.1: Communicative success and lexical alignment before introducing patterns. Population size: 10, error: 5 to 95 percentile, number of simulations: 20, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$.

Boolean decision: Does the word on place n fit the category on place n or were one or more slots skipped? Once the categories and their word order have been fully established this question should become easy to answer but as long as they are being bootstrapped a lot of inconsistent usages and hypotheses may be formed and listeners will incorrectly judge that categories have been skipped or not.

The consequence of such a misjudgement is quite severe in alignment. Remember that in alignment the listener shapes his categories towards the words filling their slots. A misjudged word, filling an incorrect slot, will push the category in the wrong direction, entrenching and eroding incorrect features. This further widens agent's misalignment and might induce an ever-increasing misalignment. Taken as such I believe the problem of optionality might be a more ominous issue for adaptive strategies because the shaping dynamics might simply shape in the wrong direction.

7.2 Extending the Weighted Adaptive Strategy

In order to further categorize a lexicon we need a population that has arrived at a lexicon that shows internal complexity. From the three types of data we have tested the Weighted Adaptive Strategy on (simulated, mushroom, robotic), the mushroom data set resulted in the largest and most diverse emergent lexicons. This data set was introduced in Chapter 5, Section 5.5.1, page 165 and consisted of 12 multi-valued features resulting in a total of 58 attributes. Each of the 8124 instances represents a mushroom for which the population develops a lexicon, categorizing the data set in the process. As explained on page 165 each attribute is created by combining a feature-name and its value. For example the attribute "cap-shape=convex" is composed from the feature "cap-shape" and value "convex". Since there are five different values for the feature cap-shape there are five related attributes.

We thus start from a population of agents that have played enough language games to establish a more or less stable lexical system that allows them to successfully communicate about the objects in their world. In Figure 7.1 we see the communicative success and lexical alignment for a population of ten agents that has, over the course of 50000 games (10000 per agent), bootstrapped a shared, stable lexicon about the mushroom data set, leading to a lexicon of some 80 words. Note that lexical alignment lies between 70 and 80% which means word meanings are not fully aligned.

7.2.1 Representational aspects

The agents have to bootstrap and align an ordered list of slots, with each slot representing a category. Each category should categorize one or more words from the lexicon. Preferably every word in the lexicon should fit into at least one of the slots. The categories themselves are also multi-dimensional, and thus show an internal structure. The strategy extends the Weighted Adaptive Strategy which means categories are represented as weighted sets.

A key difference is that we assume that the atomic elements of categories are no longer attributes but feature-names. Remember that an attribute (e.g. [odor=n]) is the combination of a feature-name (e.g. [odor]) and its value (e.g. [n]). Semantic categories refer only to (sets of) feature-names. The reasons I chose to move the categories to the level of the feature-names is that it is more in line with the example of adjectival ordering. If we look at those categories they also tend to be more general, like *color* and *size* instead of a specific value. Still the agents need to shape categories from sets of feature-names. Figure 7.2 shows an example of a pattern.

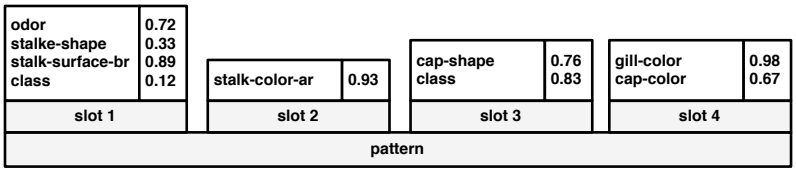


Figure 7.2: Example of category-based word order pattern. In this case only four slots are shown.

Similarity between a word meaning and a semantic category is calculated by the same similarity measure as used in the Weighted Adaptive Strategy (see equation (5.2), page 159) with the addition that an attribute (of a meaning) matches a feature-name (of a category) when it has that feature-name. The attribute [odor=n] matches with the feature-name [odor], but so does the attribute [odor=f] and all other odor-related attributes. As such a semantic category is per definition more general than a word meaning.

We want to remain true to the three basic tenets of adaptive strategies:

1. No explicit representation or enumeration of competitors
2. Flexible (similarity-based) processing
3. Internally adaptive representations that take into account feedback from flexible processing

For the representational aspects this means that an agent never represents competing word order or even competing categories. At all times there is only a single word pattern maintained by each agent. In this pattern only a single category is maintained per slot. No competing categories are represented. This is again a fairly radical approach that heavily relies on the power of the shaping dynamics that result from flexible processing and internal adaptation.

7.2.2 Production and interpretation with word order

True to adaptive strategies learning and alignment of the categories and the pattern rely on the feedback from the (flexible) linguistic processing capabilities of the agents. Therefore production and interpretation with a given pattern is explained first. We start with production.

As speakers, agents first produce in exactly the same way as described in Chapter 5, Section 5.3.2, page 159. This results in a set of words, which in the previous chapters was uttered according to the order in which the words

```
———————— Function ProduceWithPattern(topic, agent) ————————
// First produce as normal
utterance ← ProduceWeighted(topic, agent);
orderedUtterance ← nil;
pattern ← Pattern(agent);
ForEach category in pattern do
  bestWord ← Find word in utterance most similar to category;
  If Similarity(bestWord, category) > 0.0
  then add word to end of orderedUtterance;
       and remove word from utterance;
  else skip slot; // No compatible word found
End ForEach
// It is possible not all words got matched to a category
orderedUtterance ← Append(orderedUtterance,utterance);
Return orderedUtterance;
————————————————— End ProduceWithPattern —————————————————
```

Algorithm 17: Function ProduceWithPattern(topic, agent) produces an utterance ordered according to the pattern of the agent.

were added by the production algorithm. Given a pattern the agents order the words such that they fit the pattern as best as possible. The pattern is thus only used in post-processing to order the words, it does not influence which words are chosen.

For each slot in the pattern the agent looks which word matches best with the associated semantic category. This match is based on similarity and is only accepted when the similarity is over 0.0[1]. The reason for this threshold is the same as the threshold in the adaptive production algorithms from chapter 5. We need to limit the flexibility of the strategy. Without this limitation any word would in the end "fit" with a category and as a result an agent would never skip a category.

When no word with a positive similarity is found the slot is simply skipped and the next slot is tried. This simple step thus brings out the optionality of expressing a certain category and makes the task of the listener to correctly map his pattern on the utterance much harder.

It can happen that some words did not fit any category, either because the pattern simply did not have enough categories or it did not have any category similar enough to those words. In this case the pattern needs to be extended, which is covered in the next section. But as this extension will not help the agent now he simply utters all uncategorized words at the end of the utterance. Algorithm 17 shows the production process in pseudocode.

For example Figure 7.3 schematically illustrates this process for a small

[1]Remember that the similarity measure ranges from -1 to 1.

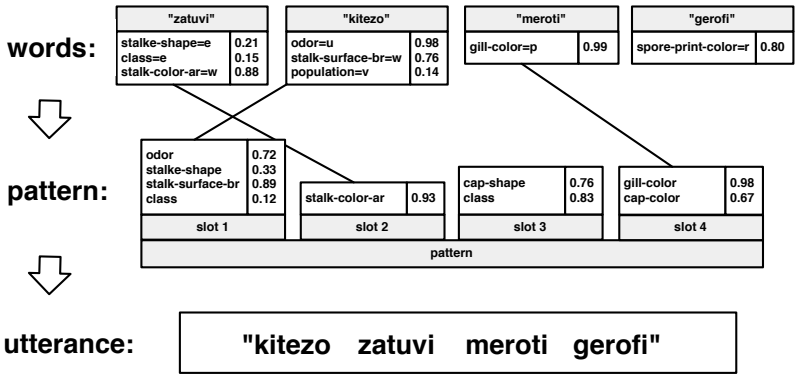


Figure 7.3: Schematic representation of production with use of a word order pattern. A pattern of four slots is used to order four words. The third slot is skipped because no word matches on the category and one word ("gerofi") cannot be categorised as part of the pattern. Given the production procedure explained in the text this results in the utterance "kitezo zatuvi meroti gerofi".

pattern of four slots and and utterance of four words. Each slot contains a semantic category represented as a weighted set of feature-names. Each word is represented by a word form and a weighted set of attributes. The production process matches "kitezo" on the first category and "zatuvi" on the second. For the third slot there is no applicable word since neither "meroti" nor "gerofi" relate to [class] and/or [cap-shape]. The word "meroti" matches with the fourth slot which results in the utterance "kitezo zatuvi meroti gerofi".

In interpretation, the listener first interprets the utterance with the interpretation function from the Weighted Adaptive Strategy. The ordering does not play a role in understanding the meaning of the utterance and thus does not influence communicative success. adding such an influence is a possibility for future extensions of the word order strategy. After interpretation the agent checks whether the word order is compatible with his own category-based word order pattern.

The listener matches the utterance to his own pattern by matching each consecutive word to the first open slot which has similarity over 0.0. For example for the first word the listener starts matching to the first slot only looking further when the similarity is negative. For the nth word the agent starts matching from the position the previous word matched on (which is at least $n-1$ if no slots were skipped so far) and again matches with the first

```
—— Function MapUtteranceToPattern(utterance, context, agent) ——
pattern ← Pattern(agent); // pattern is a deep copy
patternMatch ← nil;
ForEach word in utterance do
  ForEach category in pattern do
    pattern ← remove category from pattern;
    // the above removal does not change Pattern(agent)
    If Similarity(category, word) <= 0.0;
    then //skip the category
      add pairing <category, nil> to patternMatch;
    else add pairing <category, word> to patternMatch;
  End ForEach
  If pattern is empty
  then add all remaining words as <nil, word> to patternMatch;
       and Return patternMatch;
End ForEach
Return patternMatch;
———————————— End MapUtteranceToPattern ————————————
```

Algorithm 18: Function MapUtteranceToPattern(utterance, context, agent) maps a received utterance to the pattern of the listener.

slot that returns a non-negative similarity. As such the listener can also skip slots and handle optionality. It follows that as soon as one word cannot match on any of the remaining slots the process stops and all remaining words are considered unmatched with the word order pattern. This process is described in pseudocode by Algorithm **??**.

When the categories of speaker and listener are not aligned very well then function `MapUtteranceToPattern` might mistakenly skip categories that the speaker did not skip. In the bootstrapping phase this happens quite often and once a word has been incorrectly categorised the result is most often that all remaining words fail as well. This is the problem of optionality introduced earlier. As will become clear when discussing alignment, incorrectly matching words to categories has dire consequences for the shaping process.

What is most important is that both for the speaker and the listener the manner in which words are matched unto the pattern is flexible. It is driven by similarity and allows words to "fit" a category even when there is mismatch between the attributes of the word meaning and the feature-names of the category. So far the pattern strategy described thus holds true to the first two tenets of adaptive strategies. First there is no explicit representation of competitors, not at the level of the pattern and not at the level of the individual categories. And second the matching of words to the pattern is flexible and allows a mismatch between meaning and category.

```
———————— Function MeaningToCategory(word) ————————
category ← nil;
ForEach attribute in f_meaning(word) do
  feature-name ← extract feature-name from attribute;
  add feature-name to category with low initial certainty score;
End ForEach
Return category;
———————————— End MeaningToCategory ————————————
```

Algorithm 19: Function MeaningToCategory(word) creates an (initial) category based on the meaning of the given word.

7.2.3 Building a word order pattern

So far we have covered representation and processing. In this section the most important part, that of bootstrapping and aligning a category-based word order pattern is discussed. In order to meet the requirements of an adaptive strategy the learning and alignment needs to internally adapt the pattern and its categories based on the feedback from the flexible processing.

Bootstrapping such a pattern involves two parallel processes. One is the addition or removal of slots in the pattern and the second is the shaping of the categories of each slot. We start by discussing the addition of new slots to the pattern.

Adding new slots to the pattern

From the production procedure explained in the previous section we have learned that some words can remain unmatched with the pattern and are randomly uttered at the end (see Algorithm 17). Agents are equipped with a diagnostic which diagnoses for exactly the case in which words did not match the word order pattern[2]. In response to the diagnostic the agent picks the first unmatched word and creates a new slot at the end of the pattern consisting of all the feature-names of the attributes of that word. In the schematic representation of Figure 7.3 the word "gerofi" did not match with the pattern and thus the pattern is extended with a fifth slot and semantic category {[spore-print-color]}. A maximum of only one slot can be added per interaction resulting in an incremental built-up of the pattern and the associated semantic categories. Pseudocode is given in Algorithm 20 which relies on Algorithm 19.

As a listener, agents have a similar way of extending the word order pattern. When one or more words could not be matched against the pattern

[2]This also covers the case in which the agent has no pattern. In this case all words mismatched.

```
Function ExtendPattern(pattern, unmatchedWords)
newCategory ← MeaningToCategory(first of unmatchedWords);
add newCategory to end of pattern;
Return pattern;
End ExtendPattern
```

Algorithm 20: Function ExtendPattern(pattern, unmatchedWords) extends the given pattern with one new category based on the first unmatched word.

so far, the first non-matching word is chosen and a category for it is added to the existing pattern. As categories are *weighted* sets of feature-names, every element of a new category is given an initial low score the same way as is done for word meanings in the Weighted Adaptive Strategy (see Chapter 5, Section 5.3.3, page 161).

This covers how agents add new initial categories to their pattern. Further shaping these categories to their conventional shared form is done in much the same way as word meanings are updated by the Weighted Adaptive Strategy.

Shaping categories in the pattern

Using function `MapUtteranceToPattern` (see Algorithm 18) the listener tried to associate each word to a category in his pattern. In the successful case, in which he could match every word to a semantic category, the listener updates the categories by incrementing the score of the matched feature-names (entrenchment) and decrementing the scores of the unmatched feature-names (erosion). In the long term, semantic categories should, by following this update rule, better conform to the lexicons of the agents. This is in fact exactly the same mechanism underlying the way word meanings have been shaped, the same code (or pseudocode) can be re-used here. The only small difference is that similarity is measured between a weighted set of attributes (meaning) and a weighted set of feature-names (category).

Categories are shaped toward the meanings of the words that fill their slot. Note that the agents need not have identical word meanings, so even in case both agents are correctly shaping toward the correct word meaning, they might still not be shaping to identical sets. This problem has also been addressed in the previous chapter where it was shown that adaptive strategies show robustness against such a misalignment although it does show an impact.

But shaping towards misaligned meanings is not the biggest problem agents face. As explained earlier, function `MapUtteranceToPattern` might have made mistakes when mapping words to categories. In the case where the listener was able to match all words, this is not so likely but still possible. The result is that the categories that were incorrectly mapped to the wrong word

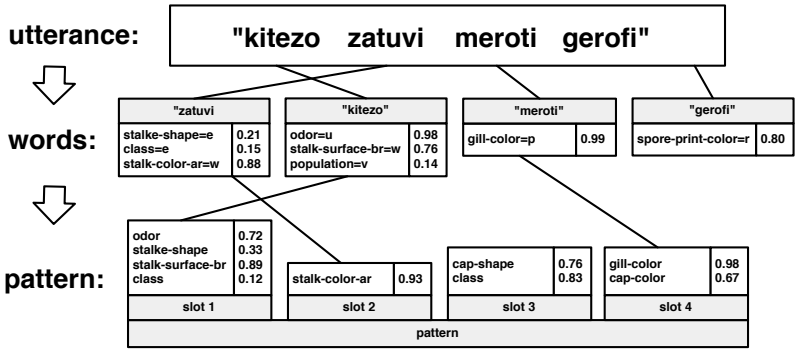


Figure 7.4: Schematic representation of interpretation with use of a word order pattern. A pattern of four slots is used to match four words of an utterance. The third slot is skipped because no word matches on the category and one word ("gerofi") cannot be categorised as part of the pattern.

are updated based on the incorrect word meaning. The category thus shapes towards the incorrect word meaning which is obviously more destructive than shaping towards a misaligned word meaning.

Although possible, in the case in which the listener could match all words, he doesn't have a good reason to believe he made any mistakes and so will just rely on the shaping dynamics to withstand potential incorrect entrenchments and erosions.

In the case that a listener could not match all the uttered words to his pattern the situation changes. His pattern was either too short or some categories are misaligned to such an extent that they lead incorrect categorizations. The listener might not always know with certainty which of the two cases is the reason and so relies on a a slightly more drastic alignment operation.

An example of this kind of uncertainty is shown in Figure 7.4. It shows an utterance as it is matched to a pattern in interpretation. The first two words successfully match on the first two semantic categories. The third word does not reach positive similarity with the third slot and is tried (successfully) with the fourth, skipping the third semantic category. The final word "gerofi" remains unmatched. As previously discussed the listener adds a new semantic category based on "gerofi", but, as slots were skipped the reason for the mismatch might lie in misaligned semantic categories. Maybe the speaker did intend for the third word "meroti" to match on the third semantic category.

The main difficulty thus lies in the optionality of slots in the pattern. The listener can never be certain when the speaker intended to skip a slot. So

```
Function AlignPattern(patternMatch, pattern, utterance)
// parameter patternMatch was calculated by MapUtteranceToPattern
If all words successfully matched in patternMatch;
then update matched categories according to weighted adaptive updating;
else
  unmatchedWords ← retrieve all unmatched words from patternMatch;
  firstSkippedCategory ← retrieve first skipped category from patternMatch;
  ExtendPattern(pattern, unmatchedWords); // defined earlier

  // Now comes the addition of features to existing categories
  Loop for category in pattern
  starting from firstSkippedCategeory do
    categoryPos ← position of category in pattern;
    potentialWord ← word in utterance at position categoryPos;
    categoryAddition ← MeaningToCategory(potentialWord);
    category ← category ∪_w categoryAddition;
  End Loop
End If
Return pattern;
End AlignPattern
```

Algorithm 21: Function AlignPattern(patternMatch, pattern, utterance) aligns the pattern based on the patternMatch created by MapUtterance-ToPattern.

whenever the listener skips a slot it might be incorrect. The reason for this lies in differences in the agents' semantic categories[3]. Next to adding a new slot the listener also extends the semantic categories of the pattern, starting from the first skipped semantic category. He adds the feature-names of the word meanings that did not (but maybe should have) matched with it. In the example this means adding, with low initial certainty scores, the feature-name [gill-color] (from "meroti") to the semantic category of the third slot and adding [spore-print-color] (from "gerofi") to the fourth slot.

Of course the first two, successfully matched, semantic categories are updated by the normal updating mechanics. For example the first word "kitezo" is used to update the first semantic category. From the four feature-names in the category only [odor] and [stalk-surface-br] matched with the meaning of "kitezo" and will be more entrenched. The other two feature-names are eroded. Just like in the Weighted Adaptive Strategy the certainty score is not directly manipulated but is a function of two underlying variables counting the number of matches and mismatches (please see Chapter 5, Section 5.3.4, page 161).

Pseudocode for the alignment is shown in Algorithm 21. Although the step that adds new features to a category might look complicated and advanced it is actually very blunt in the way it extends these existing categories and has a

[3]It could also be because of differences in the word meanings if the lexicons are highly misaligned.

high chance of adding mostly incorrect features. The idea is that the adaptive shaping dynamics should be robust enough to eliminate the incorrect features and entrench the needed required features.

To summarize the listener takes the following steps with regard to updating his pattern:

1. Apply regular update mechanics (entrenchment and erosion) to successfully matched categories.

2. When at least one word remained unmatched with a category add an new category at the end of the pattern based on the lexical meaning of the first unmatched word.

3. When, in addition, slots have been skipped (which happens very regularly) add new feature-names to the semantic categories starting from the first skipped slot.

The last step of adding new features is also required to make sure that there is no large bias towards atomic or small categories. This step can be seen as refreshing the category with new "raw" material (features) so that the shaping dynamics can try and find a better variant of the category.

7.3 Measures for categories and patterns

As explained earlier, the patterns themselves do not influence communicative success or the meanings of the words. So we no longer need to measure for lexical alignment or communicative success. The key measure we are interested in is whether the agents reach alignment about the categories they are bootstrapping and about the order for those categories. This leads to the measure of *category alignment*:

$$\text{Category alignment}(A,B) = \frac{|A \cap_w B| + |B \cap_w A|}{|A| + |B|} \tag{7.2}$$

The above measure is the same as was used for lexical alignment for the Weighed Adaptive Strategies. Category alignment is used to compare categories among different agents that hold the same slot in their respective patterns. For example the category in slot 1 of an agent's pattern is only compared to another agent's category also in slot 1. It are these categories that need to become aligned.

Another measure we are interested in is the amount of overlap between categories in a single pattern (of one agent). This measure is called *Category overlap* and is measured by measuring the $Overlap_w$ for each category in a

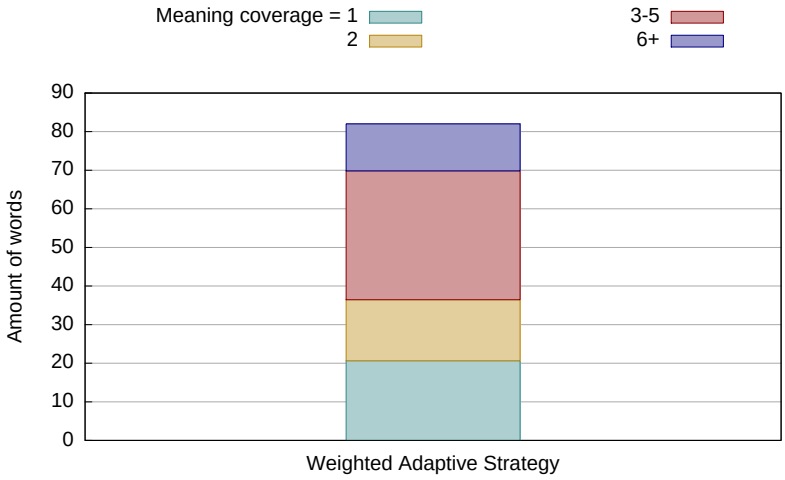


Figure 7.5: Lexicon size and meaning coverage before adding patterns. It shows that the average lexicon size is slightly over 80 words and that meanings of different schematicity have emerged. Population size: 10, number of simulations: 20, $\delta_{inh} = \delta_{dec} = 0.3$ and $\delta_{initial} = 1.0$.

pattern. $Overlap_w$ was the weighted overlap (similarity) metric used by the Weighted Adaptive Strategy and defined on page 159. I chose to use this measure because it ranges from -1 to 1 and can thus show dissimilarity better than Category alignment can.

7.4 Experimental results

As explained earlier we do not start these experiments from scratch but first allow a population of agents to bootstrap a lexicon. The results of this were shown in Figure 7.1. A population of ten agents has thus bootstrapped and aligned a lexicon to a more or less stable state. Figure 7.1 showed that the population has reached full communicative success and lexical alignment of almost 90%. The lexicon size and meaning schematicity is shown in Figure 7.5. It shows that a lexicons with an average of just over 80 words has emerged and that words cover both specific and general meanings.

There are two core robustness-related challenges that the strategy depicted above needs to tackle. The first stems from the fact that agents align their semantic categories based on their lexical entries. These lexical entries are not

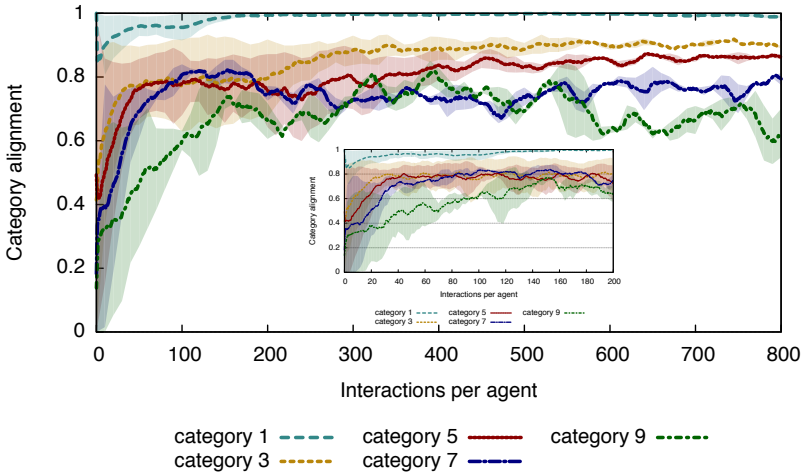


Figure 7.6: Category alignment of the first five semantic categories. For the first five slots the category alignment of each category throughout the population is plotted. Alignment gradually increases and tends to be higher the more the category is to the beginning of the pattern. Population size: 10, nr of simulations: 10, error: 10 to 90 percentile.

fully aligned themselves and thus complicate the alignment of the individual categories. The second problem arises from the optionality of the slots which leads agents to incorrectly align a category based on the wrong word.

The research question is to see to what extent, if at all, an adaptive strategy as outlined above can bootstrap and align a category-based word order pattern in a population of agents. This adaptive strategy does not have the ability to represent competing word orders or competing categories. Nevertheless a single word order (or better category order) should emerge.

Figure 7.6 shows, for the categories in slots 1, 3, 5, 7 and 9, the average alignment of that category in the population. The population gradually aligns the categories although most max out between 70 and 90% alignment. Only the first semantic category becomes more or less fully aligned. There are at least three reasons why not all categories become fully aligned. First because, just like with lexical meanings, there is no pressure to reach full alignment. If it is sufficient for the agents to order their words based on a pattern for which the categories are 70% similar then they stop making structural changes to the semantic categories (e.g. adding new feature-names). Second, the lexical meanings themselves show an alignment of only max 80%, and these meanings

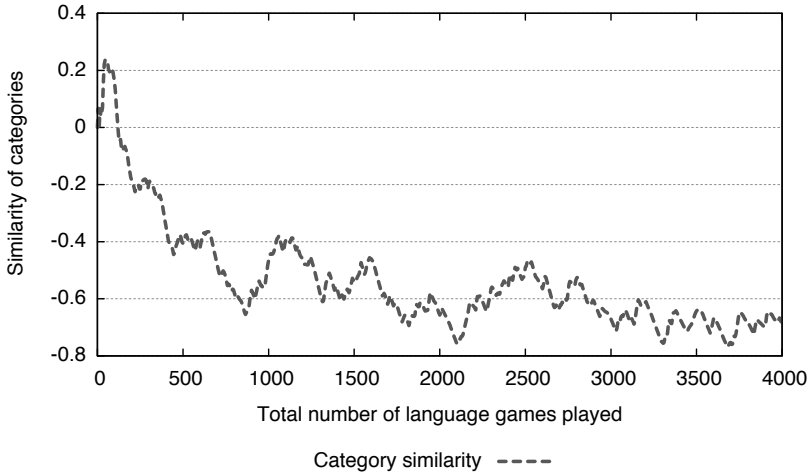


Figure 7.7: Average category overlap in a word order pattern. Over time the categories move further away from each other in the semantic space.

serve as the basis to erode and entrench the feature-names in the categories. A third reason, which also explains why categories become less aligned the further their position in the pattern, has to do with the increased uncertainty agents face when progressing in the pattern. The further a semantic category resides in the pattern the higher the chances that a slot has been skipped by the speaker and the more uncertain the hearer can correctly map a lexical meaning onto it.

Not only do the categories align throughout the population but the categories within the emergent pattern become less and less similar over time as shown in Figure 7.7. Over the course of 4000 games the average similarity of the categories decreases considerably. Categories thus move away from each other in the semantic space, covering as much of the space as possible.

7.4.1 Co-evolution of lexical and semantic categories

Because the semantic categorization in the pattern depends on and derives from the meanings captured in the lexicon it was assumed a stable lexicon first needs to be established before the pattern capabilities could successfully operate.

Surprisingly, as shown in Figure 7.8, the two systems can be bootstrapped at the same time, where semantic categories co-evolve with word meanings. It

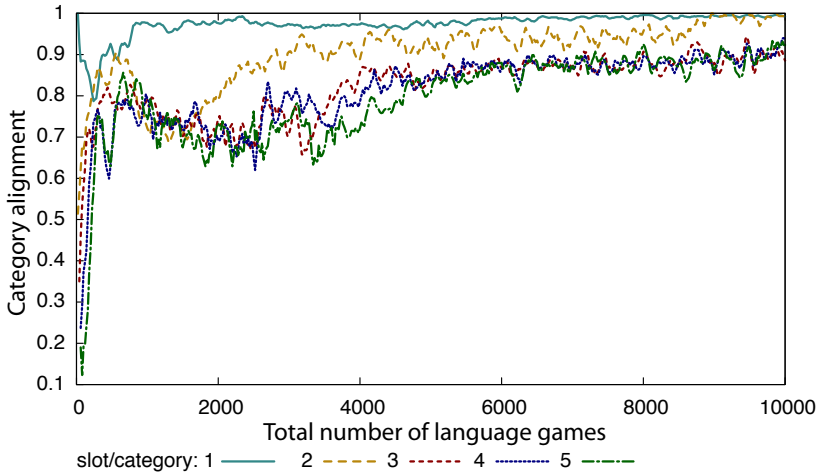


Figure 7.8: Co-evolution of lexical and semantic categories. Average category alignment of the first five semantic categories. Experimental parameters: Data set: mushroom, population size: 5.

shows that they still arrive arrive at aligned semantic categories for the word order pattern. Each pattern undergoes a dip in alignment before climbing higher. The deeper the category resides in the pattern the longer it takes to climb out of it. For example the first category enters and exits the lower level of alignment in the first 1000 games. The second category starts lowering around game 1000 and exits around game 3000. One explanation for this behaviour lies in the fact that categories that appear earlier are used more often and thus have more opportunities for adaptation. A second reason might lie in the co-evolutionary dynamics as also the words that appear near the beginning of the pattern are used more often and thus tend to stabilize faster in their meaning. As their meaning serves as the basis for the semantic category its stability is rather important.

7.5 Conclusion

The main research question addressed in this chapter was whether adaptive strategies could be extended to a non-lexical domain. We took inspiration from adjectival ordering in human languages. Adjectives are not ordered randomly but follow an ordering based on their semantic properties.

The conventionalization task was thus the bootstrapping and alignment of a word order pattern based on semantic categories. These semantic categories are derived from the lexical meanings. I discovered two sub-problems that needed to be solved in order to successfully reach such a pattern.

First the agents need to be able to align categories based on lexical meanings that might not have been fully aligned themselves. From the previous chapter we already know that adaptive strategies could bootstrap and align a lexicon when the experiences of the objects were not the same. From an abstract point of view this is the same problem as aligning categories towards non-fully aligned meanings. So we could be fairly confident that this would not pose insurmountable problems for an adaptive approach.

The second problem, which we did not encounter before, stemmed from the fact that agents could skip categories in production and interpretation. This optionality of slots resulted however in much greater uncertainty about whether or not a word was correctly matched (or not matched) to a category. In case of such a miscategorization the agent would align the category based on the incorrect word meaning. This is obviously more severe than aligning towards a slightly different word meaning (problem 1).

The adaptive approach that we proposed embodied the three basic tenets for adaptive strategies as introduced in Chapter 5, Section 5.1. For a category-based word order pattern these tenets translated as follows:

1. Agents never maintain more than a single pattern and will not ever explicitly represent competing word orders. The same holds for the categories themselves. At no point will a slot enumerate competing categories for that slot. At all times only a single (adaptive) category can be maintained.

2. Agents can flexibly process (produce and interpret) using these patterns. A word can still be categorised to a slot even though its meaning is not fully compliant with the containing category. This process gives information about which parts of the category were beneficial for the categorisation and which were not.

3. In alignment each feature of each used category is updated according to the feedback from flexible processing. Here we went again for a weighted variant, which meant that certainty scores for feature were increased and decreased. Features could also be removed if their certainty dropped below zero and features could be added in an exceptional case.

The results have shown that both problems can be surmounted by the adaptive approach. The agents even did so in a fairly limited amount of games, much more limited than it took the lexicon to become established. This again

shows the remarkable robustness of adaptive approaches. The alignment did not for all categories reach 100%. In fact only the very first category achieved such a status. But just as in lexical experiments, full alignment is often not necessary.

Categories that were located more to the end of the pattern did take longer to become aligned and overall became slightly less aligned. One reason for this is that they are simply used less and can only become established once the previous slots have been added. A second reason is that in interpretation the further the agent progresses in his match of the utterance on his pattern the higher the chance that he might have miscategorised a category. Once a category has been miscategorised this is quite devastating for the remained of the utterance to be matched successfully. In short, the problem of optionality increases the further the listener progresses in a pattern.

Another observation in the experimental results was that the category overlap in a pattern of a single agent decreased over the course of time. So categories became less and less similar inside a pattern. For example, over time the category of slot 1 became less and less similar to the categories in slots 2 to n. In fact, the categories became almost disjoint. This emergent feature is interesting and beneficial for two reasons. First, this leads to less ambiguity because when categories are dissimilar, a single word is less likely to be categorisable by multiple categories. Second it also means that the categories quite optimally cover the space of lexical meanings, partitioning this space into almost disjoint regions.

What is especially interesting about this phenomenon is that nowhere in the adaptive algorithm is there any explicit pressure for dissimilarity among the categories. This is an emergent effect from the local adaptations by the adaptive strategy.

Finally I have also shown that it was not even necessary to first bootstrap a stable lexicon and then in a second phase bootstrap a word order pattern. It turned out that the word order pattern could be bootstrapped at the same time as the lexicon was bootstrapped. This is a highly non-trivial feat since the categories are based on the word meanings, but these word meanings themselves are still undergoing heavy chance. The results showed that the categories only started to become more and more aligned after a few thousand games.

Chapter 8

Conclusions and future work

In this thesis I have introduced different language games, a range of conventionalization problems and for each of those multiple language strategies. In the individual conclusions of each chapter I have already highlighted the most important findings of each chapter. I will not repeat those here, instead I want to focus on some of the things that I have tried to put forward in the thesis but could not easily be included at any given point, or at least not in full.

Some arguments can only be made when combining the insights from more than one chapter or after both competitive and adaptive strategies have been introduced. These points that I have tried to make at different points throughout the thesis might have been lost in between pseudocode and bar charts. I believe that here, in the concluding chapter, is the best place to put them back in the spotlight as I believe some are more important than all the graphs and pseudocode of this thesis combined.

In the final section I also give some pointers for future research.

8.1 Putting conventionalization problems first

A more methodological contribution that I have incorporated in this thesis is to change the manner in which language games have been traditionally classified. Traditionally a language game incorporated the interaction script, the processing and learning capabilities of the agents and more often than not left the conventionalization problem or problems unspecified. For example Steels & Belpaeme (2005) defines a Guessing Game based on the interaction script. Likewise Vogt & Coumans (2003) defines three games, the observational game, the guessing game and the selfish game by their interaction script. To each of these games a particular type of processing and learning is associated.

Starting from around 2005, I believe this type of definition has become less

popular. Van Looveren (2005) already moves away from giving the interaction script its central role, although a strategy (processing and learning) is still part of a game. De Vylder (2007) clearly does no longer rely on the interaction script as the definitional feature. I believe de Vylder's thesis could be seen as one of the first attempts to move the notion of a conventionalization problem more to the foreground.

In this thesis I further expand the approach pioneered by De Vylder (2007) and separate a language game from a particular type of processing and learning and from its interaction script. I argue that it is better to put the notion of a *conventionalization problem* as the definitional feature of a language game. The processing, learning, and even the representational capabilities of the agents are not part of the game but are part of a language strategy. Examples of this new approach are found in each chapter of the thesis, an overview is given in Table 8.1.

As an example, in Chapter 2 I define a Naming Game through the problem of *word form competition*. I discuss that this problem can be evoked through different types of scripts. In the script I use, I still rely (for reasons explained there) on notions such as a context and even pointing. A simpler script is that of the observational game introduced by Vogt & Coumans (2003) or Baronchelli (2012).

Talking about the observational game by Vogt & Coumans (2003) in the classification I propose both his observational game and his guessing game only evoke the problem of word form competition, although they have different scripts. In this thesis they would thus both be categorized as Naming Games, even though their scripts are different. The selfish game would be considered a Minimal Guessing Game since it evokes mapping uncertainty (and word form competition).

This sort of categorization is much harder to apply to embodied language games. For example the Grounded Naming Game (Steels & Loetzsch, 2012) deals with problems way beyond word form competition. It has to deal with the problem of establishing identity for objects and these objects show internal complexity, even operating in a continuous visual space. In the scheme I suggest this game would thus not be a Naming Game. I don't see this as a particularly large problem since grounded language games are overall concerned with very different kinds of problems, mostly stemming from the embodiment.

8.2 Comparison, verification and reimplementation of existing strategies

Chapters two to four do not primarily focus on novel original strategies. They are intended to introduce the necessary concepts and show the reader what sort of problems have been tackled by non-grounded language games.

By re-implementing these strategies I was able to verify and sometimes further illuminate many results that have been previously published. For example in Chapter 3 I am the first to re-implement the model by De Beule *et al.* (2006).

By comparing the results of the different strategies, something that for most strategies was not done before, I also could uncover new insights not available before. For example in Chapter 2 the addition of a scoring mechanism only slightly improves upon the Minimal NG Strategy which does not implement a scoring scheme. When combined with the problem of mapping uncertainty in Chapter 3 it turned out that a Naming Game strategy with a scoring scheme did show a dramatic impact on the results.

I have in Chapter 3 expanded heavily on the fact that the word from competition problem is still at play in a Minimal Guessing Game. I have shown that this problem might have been overlooked by modelers which only focused on cross-situational strategies. This might have resulted in them undervaluing cross-situational learning since the resultant lexical systems are polluted by massive amount of form competitors. It also takes much longer to achieve a stable lexicon because many more words need to become aligned.

I have also shown that not all types of cross-situational learning can successfully tackle the mapping uncertainty problem. The set-based cross-situational strategy was unable to scale towards large populations and large meaning spaces because it was not robust against inconsistent input. In my findings, only cross-situational strategies that track the cross-situational statistics with a scoring mechanism (such as the Frequency CS and the Flexible Strategy) show enough robustness to scale up in a multi-agent bootstrapping context.

What is new in Chapter 4 is that I stress the importance of usage-based properties in the emergent lexical systems and the importance of cognitive grammar and later (in Chapter 6) even grammaticalization for models of word meaning. I have shown that approaches that add biases which reduce the initial hypothesis space considerably are bound to severely bias the possible emergent lexicons. Most likely this bias will be toward highly atomic lexicons where meanings are disjoint, show no or minimal schematicity (general and specific meanings) or usage-based properties. The Discriminative CS Strategy was an example of such an approach.

8.3 Computational models for two views on word learning

What in my opinion has been severely lacking from computational modeling approaches to the emergence of lexical systems is the acknowledgement of two competing views on what lexical acquisition entails. Most often, if not always, a model is embedded in a particular paradigm but this is not made explicit.

Understanding that two views exist makes interpretation of computational models much more straightforward. The first and most popular view sees lexical acquisition as a mapping problem. In this view meanings are equated to concepts and it is assumed they are primarily established non linguistically and before lexical acquisition starts (Bloom, 2000). A learner thus has available most if not all concepts and all that remains to be done is "map" new words to those meanings.

In this mapping approach, competitive cross-situational models make a lot of sense. Siskind (1996) explicitly defined the acquisition problem as a mapping problem when he first operationalised cross-situational learning. In fact even in the title he speaks of "learning word-to-meaning mappings". I believe however that in some cases this underlying view of lexical acquisition is forgotten and it was assumed that competitive cross-situational learning could also tackle the problem I call meaning uncertainty. This was for me the main motivation to include Chapter 4, to show the limits of competitive approaches and re-iterate their underlying assumptions.

The second view of lexical acquisition sees learning more as a constructivist process. Word meanings are formed and shaped in a usage-based manner with a direct impact from interaction and communication itself. Obviously this is a more difficult task since it is not assumed that a learner can rely on pre-established meanings. This view leads to the formulation of the problem of meaning uncertainty in Chapter 4 and finally to the adaptive strategies from Chapter 5.

The adaptive strategies I propose subscribe to this second view of lexical acquisition and are heavily inspired by both cognitive grammar (Langacker, 1987, 1991b) and grammaticalization theory (Traugott & Heine, 1991; Heine & Kuteva, 2007). In this second view all kinds of interesting phenomena such as slow mapping (Carey & Bartlett, 1978), extension, and even re-analysis (Timberlake, 1977; Hoefler & Smith, 2008) and metaphor (Boroditsky, 2000) can started to be investigated. I believe the adaptive strategies I propose can be taken as a good starting point for such further investigations.

8.4 The link between lexicon and grammar

The main three influences on my thinking about language all propose an intimate link between lexicon and grammar and see semantics and meaning as a fundamental part of both.

Construction Grammar: Construction grammar posits that all linguistic knowledge, both lexical and grammatical, is represented by the same fundamental building block, a construction. Lexical and grammatical constructions lie on a continuum with at no point a radical breaking point. Every construction contains both meaning and form, no matter how concrete or abstract.

Cognitive Grammar: Cognitive grammar stresses the exact same point as construction grammar. Langacker (2000a) speaks even of the rule/list fallacy which sees linguistic knowledge as a separate list of lexical entries and a set of (grammatical) rules.

Grammaticalization Theory: In grammaticalization theory, grammatical constructions are assumed to be shaped out of lexical items. These changes are in first instance based on semantics and only later phonological and syntactical changes can occur. Grammaticalization theory thus presupposes that both lexical and grammatical knowledge must lie in the same linguistic space because lexical items can gradually transform in grammatical items.

These influences can also explain why I believe that any theory of the emergence of lexical systems needs to take semantics and meaning very seriously. If word semantics are even important for the shaping of grammar then it would seem rather odd if they could be seen as peripheral in the emergence of lexical entries.

I included Chapter 7 to show that adaptive strategies can be applied both in the lexical and in the grammatical domain. Grammaticalization is in its essence also an adaptive strategy. Obviously a much more advanced adaptive strategy compared to the ones I have proposed in this thesis, but it embodies the core principles of adaptive strategies. I find it much more difficult to envision competitive strategies convincingly scaling to grammatical problems.

8.5 Directions for future research

There are quite a few interesting avenues for future research, both in the lexical as in the grammatical domain.

8.5.1 The combination of competitive and adaptive strategies

In this thesis I have drawn a distinction between competitive and adaptive strategies. Competitive strategies enumerate competing hypotheses, whereas adaptive strategies internally adapt a single hypothesis.

The two approaches are, however, not incompatible and it is possible to combine them into a single strategy. What I imagine is a (restricted) enumeration of competing hypotheses, where each of these hypotheses can be shaped by an adaptive approach.

There already exist examples where adaptive and competitive alignment mechanisms have been combined. For example Steels & Loetzsch (2012) implement the Grounded Naming Game by combining adaptive strategies in the continuous domain and competitive strategies in the symbolic domain. The same is done by Bleys (2012) who employs an adaptive strategy in the continuous domain of color and then applies competitive dynamics to implement a selectionist competition between strategies. What these approaches have in common is that the adaptive strategy operates in the continuous domain and the competitive in the symbolic.

Adaptive learning indeed feels more intuitive when learning takes place in a continuous space. Indeed many models operating in a continuous domain employ adaptive representations and flexible processing. For example de Boer (2001) presents an adaptive model to explain the origins of vowel systems, not surprisingly in a continuous space. Most models proposed to explain the emergence of color terms, for which the meanings are represented in a three-dimensional continuous space, are also adaptive in nature (Belpaeme & Bleys, 2005; Bleys, 2010; Belpaeme, 2002).

The earlier lexical work with embodied agents by Vogt also combines adaptive and competitive learning. For example in (Vogt, 2001, 2002, 2003) agents first play discrimination games employing an adaptive strategy. These discrimination games are not peer-to-peer but are played by each agent on his own in which the agents create discriminating categories for their environment. Categories are represented as prototypes (feature vectors) and are adapted as to better discriminate objects. In a later stage the agents then express these categories but no longer adapt the internals of the categories. The lexical learning, which is now a mapping task, is solved in a more or less competitive cross-situational manner.

What I have shown in this thesis is that adaptive strategies can also operate in symbolic or discrete domains. In the above examples I think the result can also be achieved with just an adaptive strategy alone. De Beule & Bleys (2010) has shown that the problem addressed by Bleys (2012) can be

solved only using an adaptive strategy.

What I then intend is a more fruitful combination where the competitive aspect truly solves a problem that could not be solved adaptively. The bootstrapping of grammar, especially in the formation of syntactic categories, might provide opportunities for combinations.

8.5.2 Modeling of multiple word senses

Throughout this whole thesis I have only considered a single meaning per word. Note that this was also true for the competitive approaches. Although they associate multiple *competing* meanings to a single word, they are competing in a selectionist setting so that only a single meaning per word "survives" in the end.

In natural languages we see that almost all words have multiple related sense (Miller, 1995). Open any dictionary and this becomes obvious. These different senses are not convincingly captured by fuzzy boundaries (which meanings in adaptive strategies have), but instead must be represented as distinct entries. This introduces, I believe, a new conventionalization problem which has received quite some attention from corpus-based computational linguistics (Pantel & Lin, 2002; Yarowsky, 1995) but surprisingly little in the modeling of acquisition or emergence of lexicons.

The key problem from a usage-based perspective, as I see it, is in deciding when an extended use constitutes a new sense. I think the best starting point is an adaptive strategy but it will require non-trivial extensions to deal with this problem. Another problem is in finding semantic data rich enough to allow for multiple senses.

8.5.3 Adaptive strategies in the emergence of grammar

Computational investigations in grammar are dominated by competitive approaches (Steels & Wellens, 2006; Steels *et al.*, 2007; De Beule & K. Bergen, 2006). The main reason for this is two-fold. First, grammatical representations tend to be highly symbolic (e.g. feature structures) and in the symbolic domain, competitive approaches seem more natural to many computational linguists. Second, competitive approaches are much easier to operationalize than adaptive approaches, because of the flexible processing required by the latter.

Grammatical formalisms often rely on matching algorithms on complex, hierarchical feature structures (e.g. pattern matching, unification, merging) (Pollard & Sag, 1994; Steels, 2011b; Bergen & Chang, 2005). These algorithms

are fairly Boolean in their output, either the two structures match or unify or they do not. A Boolean output is not exactly the most flexible application process and without flexible processing it is impossible to apply adaptive alignment.

It would be interesting if a grammatical formalism (e.g. Fluid Construction Grammar) could be extended with a flexible matching algorithm which would highlight parts that mismatch and parts that match so that adaptive alignment could be achieved on complex feature structures. This would open the door for very fruitful research in the computational modeling of grammaticalization.

	Naming Game	Minimal GG
(2) word form competition	yes	yes
(3) mapping uncertainty	no	yes
(4) meaning uncertainty	no	no
(6) meaning unc + misaligned input	no	no
(7) meaning unc + misaligned input + optionality	no	no
	Compositional GG	CGG + diff. views
(2) word form competition	yes	yes
(3) mapping uncertainty	N/A	N/A
(4) meaning uncertainty	yes	yes
(6) meaning unc + misaligned input	no	yes
(7) meaning unc + misaligned input + optionality	no	no
	CGG + word-order pattern	
(2) word form competition	yes	
(3) mapping uncertainty	N/A	
(4) meaning uncertainty	yes	
(6) meaning unc + misaligned input	yes	
(7) meaning unc + misaligned input + optionality	yes	

Table 8.1: Overview of Language Games and conventionalization problems they evoke. The bracketed number in front of each conventionalization problem refers to the chapter that introduced that problem. The cells state whether or not that language game evokes the conventionalization problem (rows).

Index

References

Amari, S.i. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, **27**, 77–87, 10.1007/BF00337259. 48

Baldwin, D.A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, **62**, 875–890. 134

Barlow, M. & Kemmer, S., eds. (2000). *Usage-Based Models of Language*. Chicago University Press, Chicago. 15

Baronchelli, A. (2012). The minimal naming game: a complex systems approach. In L. Steels, ed., *Experiments in Cultural Language Evolution*, John Benjamins, Amsterdam. 46, 222

Baronchelli, A., Caglioti, E., Loreto, V. & Steels, L. (2006a). Complex systems approach to language games. In *Proceedings of the Second European Conference on Complex Systems ECCS'06*, Oxford (UK). 21, 46

Baronchelli, A., Felici, M., Loreto, V., Caglioti, E. & Steels, L. (2006b). Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics*, **P06014**. 35, 46, 47

Baronchelli, A., DallAsta, L., Barrat, A. & Loreto, V. (2007). The role of topology on the dynamics of the naming game. *The European Physical Journal Special Topics*, **143**, 233–235. 36

Baronchelli, A., Loreto, V. & Steels, L. (2008). In-depth analysis of the naming game dynamics: The homogeneous mixing case. *International Journal of Modern Physics C*, **19**, 785–812. 35

Baronchelli, A., Castellano, C. & Pastor-Satorras, R. (2011). Voter models on weighted networks. *Physical Review E.*, **83, 066117**. 21, 44

Beckner, C., Blythe, R., Bybee, J., Christiansen, M.H., Croft, W., Ellis, N.C., Holland, J., Ke, J., Larsen-Freeman, D. &

Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, **59**, 1–26. 4

Belpaeme, T. (2002). *Factors Influencing the Origins of Colour Categories*. Ph.D. thesis, Vrije Universiteit Brussel. 12, 128, 226

Belpaeme, T. & Bleys, J. (2005). Explaining universal color categories through a constrained acquisition process. *Adaptive Behavior*, **13**, 293–310. 12, 226

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. 227

Beuls, K., Steels, L. & Höfer, S. (2012a). The emergence of internal agreement systems. In L. Steels, ed., *Experiments in Cultural Language Evolution*, John Benjamins, Amsterdam. 12

Beuls, K., van Trijp, R. & Wellens, P. (2012b). Diagnostics and repairs in Fluid Construction Grammar. In *Language Grounding in Robots*. 8

Blair, M. & Somerville, S.C. (2009). The importance of differentiation in young children's acquisition of expertise. *Cognition*, **112**, 259–280. 134

Bleys, J. (2010). *Language Strategies for the Domain of Colour*. Ph.D. thesis, Vrije Universiteit Brussel. 12, 226

Bleys, J. (2012). Language strategies for color. In L. Steels, ed., *Experiments in Cultural Language Evolution*, John Benjamins, Amsterdam. 12, 226

Bloom, P. (2000). *How Children Learn the Meanings of Words*. MIT Press, Cambridge, MA. 18, 61, 139, 224

Bloom, P. (2001). Roots of word learning. In M. Bowerman & S.C. Levinson, eds., *Language Acquisition and Conceptual Development*, 159–181, Cambridge University Press, Cambridge. 139

Blythe, R.A. (2009). Generic modes of consensus formation in stochastic language dynamics. *J. Stat. Mech*, P02059. 46

Blythe, R.A., Smith, K. & Smith, A.D.M. (2010). Learning times for large lexicons through cross-situational learning. *Cognitive Science*, **34**, 620–642. 22, 83, 85, 88, 98

Bod, R., Hay, J. & Jannedy, S. (2000). *Probabilistic linguistics*. MIT Press. 51

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, **75**, 1–28. 224

Bowerman, M. & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S.C. Levinson, eds., *Language Acquisition and Conceptual Development*, 132–158, Cambridge University Press, Cambridge. 18, 140

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, **8**, 25–54. 17, 103, 105

Brighton, H., Kirby, S. & Smith, K. (2005). 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. 3

Briscoe, E.J. (1998). Language as a complex adaptive system: Coevolution of language and of the language acquisition device. In H. van Halteren et al., ed., *Proceedings of Eighth Computational Linguistics in the Netherlands Conference*. 4

Bybee, J. (1998). The emergent lexicon. *CLS 34: The panels*, 421–435. 51

Carey, S. (1978). The child as word learner. In E. Wanner, M. Maratsos, M. Halle, J. Bresnan & G. Miller, eds., *Linguistic Theory and Psychological Reality*, 264–293, MIT Press, Cambridge, Ma. 61

Carey, S. & Bartlett, E. (1978). Acquiring a single new word. *Papers and reports on child language development*, **15**, 17–29. 61, 157, 224

Castellano, C., Fortunato, S. & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, **81**, 591–646. 46

Chafe, W.L., ed. (1980). *The Pear Stories*. Ablex, New York. 5

Childers, J.B. (2008). The structural alignment and comparison of events in verb acquisition. 175

Childers, J.B. & Paik, J.H. (2009). Korean- and english-speaking children use cross-situational information to learn novel predicate terms. *Journal of Child Language*, **36**, 201–224. 175

Childers, J.B. & Tomasello, M. (2001). The role of pronouns in young children's acquisition of the english transitive construction. *Developmental Psychology*, **37**, 739–748. 175

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA. 4, 14

Chomsky, N. (1980). *Rules and Representations*. Basil Blackwell, Oxford. 14

Chomsky, N. (1995). *The Minimalist Program*. MIT Press, Cambridge, MA. 14

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. 14

Clancy, P.M. (1980). *Referential Choice in English and Japanese Narrative Discourse*. Ablex, Norwood, New Jersey. 6

Cornish, H., Tamariz, M. & Kirby, S. (2009). Complex adaptive systems and the origins of adaptive structure: What experiments can tell us. *Language Learning*, **59**, 187–205. 17

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. 201

Croft, W. & Cruse, D.A. (2004). *Cognitive Linguistics*. Cambridge Textbooks in Linguistics, Cambridge University Press, Cambridge. 14, 15

Csikszentmihalyi, M. (1990). *Flow. The Psychology of Optimal Experience*. Harper and Row, New York. 28

De Beule, J. (2007). *Compositionality, Hierarchy and Recursion in Language. A Case Study in Fluid Construction Grammar*. Ph.D. thesis, Vrije Universiteit Brussel. 106

De Beule, J. & Bleys, J. (2010). Self-organization and emergence in languages: a case study for color. In A.D.M. Smith, M. Schouwstra, B. de Boer & K. Smith, eds., *The Evolution of Language: Evolang 8*, 83–91. 226

De Beule, J. & De Vylder, B. (2005). Does language shape the way we conceptualize the world? In B.G. Bara, L. Barsalou & M. Bucciarelli, eds., *the XVII Annual Conference of the Cognitive Science Society*, 559–564. 21, 51, 128

De Beule, J. & K. Bergen, B. (2006). On the emergence of compositionality. In A. Cangelosi, A. Smith & K. Smith, eds., *Proceedings of the 6th International Conference on the Evolution of Language*, 35–42, World Scientific Publishing, London. 103, 105, 227

De Beule, J., De Vylder, B. & Belpaeme, T. (2006). A cross-situational learning algorithm for damping homonymy in the guessing game. In L.M. Rocha, L.S. Yaeger, M.A. Bedau, D. Floreano, R.L. Goldstone & A. Vespignani, eds., *Artificial Life X: Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems*, 466–472, MIT Press, Cambridge, MA. 23, 33, 66, 71, 94, 95, 96, 98, 106, 107, 223

de Boer, B. (2000). Self organization in vowel systems. *Journal of Phonology*, **28**, 441–465. 37

de Boer, B. (2001). *The origins of vowel systems*. Oxford University Press, Oxford. 226

de Boer, B. & Fitch, W.T. (2010). Computer models of vocal tract evolution: an overview and critique. *Adaptive Behaviour*, **18**, 36–47. 37

De Vylder, B. (2007). *The Evolution of Conventions in Multi-Agent Systems*. Ph.D. thesis, Vrije Universiteit Brussel, Brussels. 22, 54, 82, 94, 98, 222

De Vylder, B. & Tuyls, K. (2006). How to reach linguistic consensus: A proof of convergence for the naming game. *Journal of Theoretical Biology*, **242**, 818–831. 35

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, **17**, 463–489. 6

Divina, F. & Vogt, P. (2006). A hybrid model for learning word-meaning mappings. In P. Vogt & et al., eds., *Symbol Grounding and Beyond: Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication*, 1–15, Springer Berlin/Heidelberg. 68

Downing, P. (1980). *Factors Influencing Lexical Choice in Narrative*. Ablex, Norwood, New Jersey. 6

Evans, N. & Levinson, S.C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, **32**, 429–492. 16

Fazly, A., Alishahi, A. & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science: A Multidisciplinary Journal*, **34**, 1017–1063. 99

Ferber, J. (1999). *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*. Addison Wesley Longman, Harlow, UK. 8

Ferrer-I-Cancho, R. & Forns, N. (2010). The Self-Organization of Genomes. *Complexity*, **15**, 34–36. 16

Fillmore, C.J. (1977). Scenes-and-frames semantics. In A. Zampolli, ed., *Linguistic structures processing*, 55–81, North Holland Publishing, Amsterdam. 141

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. 201

Flavell, J.H. (1992). Perspectives on perspective taking. In H. Beilin & P.B. Pufall, eds., *Piaget's Theory: Prospects and Possibilities*, 107–139, Erlbaum, Hillsdale, NJ. 203

Frank, A. & Asuncion, A. (2010). UCI machine learning repository. 117, 167

Fu, F. & Wang, L. (2008). Coevolutionary dynamics of opinions and networks: From diversity to uniformity. *Physical Review E*, **78**. 46

Fujita, M., Kuroki, Y., Ishida, T. & Doi, T.T. (2003). Autonomous behavior control architecture of entertainment humanoid robot sdr-4x. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)*, 960–967, Las Vegas, Nevada. 179

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, **29**, 737–767. 17, 56

Galantucci, B. & Garrod, S. (2011). Experimental semiotics: a review. *Frontiers in Human Neuroscience*, **5**. 17

Galantucci, B. & Steels, L. (2008). *The emergence of embodied communication in artificial agents and humans.*, 229–256. Oxford University Press. 17

Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought*. A Bradford Book, MIT Press. 145, 158

Gärdenfors, P. (2004). Cooperation and the evolution of symbolic communication. In D.K. Oller & U. Griebel, eds., *Evolution of Communication Systems: A Comparative Approach*, 237–256, MIT Press, Cambridge, MA. 158

Garrod, S. & Pickering, M.J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, **1**, 292–304. 17

Gentner, D. & Medina, J. (1998). Similarity and the development of rules. *Cognition*, 263–297. 175

Gentner, D. & Namy, L.L. (1999). Comparison in the development of categories. *Cognitive Development*, **14**, 487–513. 175

Gentner, D. & Namy, L.L. (2004). *The role of comparison in children's early word learning*, 533–568. MIT Press, Cambridge, MA. 175

Gentner, D. & Rattermann, M.J. (1991). *Language and the career of similarity*, 225–277. Cambridge University Press, London. 145, 194

Gentner, D., Loewenstein, J. & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, **95**, 393–408. 175

Gerasymova, K., Steels, L. & van Trijp, R. (2009). Aspectual morphology of russian verbs in Fluid Construction Grammar. In N. Taatgen & H. van Rijn, eds., *Proceedings of the 31th Annual Conference of the Cognitive Science Society*, Cognitive Science Society. 12

Gillette, J., Gleitman, H., Gleitman, L. & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, **73**, 135–176. 62, 68, 72

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, **1**, 3–55. 62, 139

Goldberg, A.E. (1995). *A Construction Grammar Approach to Argument Structure*. Chicago UP, Chicago. 201

Goldberg, A.E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Science*, **7**, 219–224. 14

Graham, S., Namy, L., Gentner, D. & Meagher, K. (2010). The role of comparison in preschoolers' novel object categorization. *Journal of Experimental Child Psychology*, **107**, 280–290. 134

Hammer, R. & Diesendruck, G. (2005). The role of dimensional distinctiveness in children and adults' artifact categorization. *Psychological Science*, **16**, 137–144. 134

Hartline, H., Wagner, H. & Ratliff, F. (1956). Inhibition in the eye of limulus. *J. Gen. Physiology*, **39**, 651. 48

Haspelmath, M., Dryer, M., Gil, D. & Comrie, B., eds. (2005). *The World Atlas of Language Structures*. Oxford University Press, Oxford. 141

Heine, B. & Kuteva, T. (2007). *The Genesis of Grammar: A Reconstruction*. Oxford University Press. 15, 194, 224

Heine, B., Claudi, U. & Hünnemeyer, F. (1991). *Grammaticalization: A Conceptual Framework*. University of Chicago Press, Chicago. 15, 194

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. 224

Hopper, P. (1991). Emergent grammar. In E. Traugott & B. Heine, eds., *Approaches to Grammaticalization*, John Benjamins, Amsterdam. 194

Hopper, P. & Traugott, E. (1993). *Grammaticalization*. Cambridge University Press, Cambridge. 15

Horst, J. & Samuelson, L. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, **13**, 128–157. 61

Hovland, C.I. & Hunt, E.B. (1961). Programming a model of human concept formulation. In *afips Proceedings of the Western Joint Computer Conference*, 145. 65

Iba, W., Wogulis, J. & Langley, P. (1988). Trading off simplicity and coverage in incremental concept learning. In *Proceedings of the 5th International Conference on Machine Learning*, 73–79, Morgan Kaufmann, Ann Arbor, Michigan. 167

Jaccard, P. (1901). Distribution de la flore alpine dans le bassin des dranses et dans quelques rgions voisines. *Bulletin de la Socit Vaudoise des Sciences Naturelles*, **37**, 241–272. 115

Ketteler, D., Kastrau, F., Vohn, R. & Huber, W. (2008). The subcortical role of language processing. high level linguistic features such as ambiguity-resolution and the human brain; an fmri study. *NeuroImage*, **39**, 2002 – 2009. 9

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. 103, 105

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. 17

Kirby, S. (2002). Natural language from artificial life. *Artificial Life*, **8**, 185–215. 10

Kirby, S., Smith, K. & Brighton, H. (2004). From UG to universals: Linguistic adaptation through iterated learning. *Studies in Language*, **28**, 587–607. 17

Kirby, S., Cornish, H. & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, **105**, 10681–10686. 3

Landau, B., Smith, L.B. & Jones, S.S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, **3**, 299–321. 62

Langacker, R.W. (1987). *Foundations of Cognitive Grammar. Volume 1*. Stanford University Press, Stanford. 15, 224

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. 224

Langacker, R.W. (2000a). A dynamic usage-based model. In M. Barlow & S. Kemmer, eds., *Usage-Based Models of Language*, 1–63, Chicago University Press, Chicago. 15, 161, 163, 171, 175, 225

Langacker, R.W. (2000b). *Grammar and Conceptualization*. Mouton de Gruyter, Den Haag. 117

Lenaerts, T., Jansen, B., Tuyls, K. & De Vylder, B. (2005). The evolutionary language game: An orthogonal approach. *Journal of Theoretical Biology*, **235**, 566 – 582. 35

Levinson, S.C. (2001). Language and mind: Let's get the issues straight! In M. Bowerman & S.C. Levinson, eds., *Language Acquisition and Conceptual Development*, 25–46, Cambridge University Press, Cambridge. 141

Lipowski, A. & Lipowska, D. (2008). Bio-linguistic transition and baldwin effect in an evolutionary naming-game model. *Int. J. of Mod. Phys. C*, **19**, 399 – 408. 35

Liu, R., Jia, C., Yang, H. & Wang, B. (2009). Naming game on small-world networks with geographical effects. *Physica A*. 35

Loetzsch, M. & Spranger, M. (2010). Why robots? In *The Evolution of Language (EVOLANG 8)*, 222–229, World Scientific, Singapore. 179

Loetzsch, M., van Trijp, R., & Steels, L. (2008a). 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. 128, 192

Loetzsch, M., Wellens, P., De Beule, J., Bleys, J. & van Trijp, R. (2008b). The babel2 manual. Tech. Rep. AI-Memo 01-08, AI-Lab VUB, Brussels. 8

Loewenstein, J. & Gentner, D. (2001). Spatial mapping in preschoolers: Close comparisons facilitate far mappings. *Journal of Cognition and Development*, **2**, 189–219. 175

Lu, Q., Korniss, G. & Szymanski, B. (2008). Naming games in twodimensional and small-world-connected random geometric networks. *Physical Review E*, **77**. 35

Madole, K.L. & Oakes, L.M. (1999). Making sense of infant categorization: Stable processes and changing representations. *Developmental Review*, 263–296. 135

Markman, E.M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In M. Gunnar & M. Maratsos, eds., *Modularity and Constraints in Language and Cognition: The Minnesota Symposium on Child Psychology*, vol. 25, 59–101, Erlbaum, Hillsdale, NJ. 62, 139

Markman, E.M., Wason, J.L. & Hansen, M.B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, **47**, 241–275. 112, 147

Maynard-Smith, J. & Szathmáry, E. (1995). *The Major Transitions in Evolution*. Freeman. 16

McMahon, A.M.S. (1994). *Understanding Language Change*. Cambridge University Press. 4

Miller, G.A. (1995). Wordnet: a lexical database for english. *Commun. ACM*, **38**, 39–41. 227

Mitchell, T. (1977). Version spaces: A candidate elimination approach to rule learning. In *Fifth International Joint Conference on AI*, 305–310, MIT Press, Cambridge, MA. 82

Moore, C. & Dunham, P.J., eds. (1995). *Joint Attention: Its Origins and Role in Development*. Lawrence Erlbaum Associates, Hillsdale, NJ. 6

Namy, L.L. & Gentner, D. (2002). Making a silk purse out of sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General*, **131**, 5–15. 175

Nappa, R., Wessell, A., McEldoon, K., Gleitman, L. & Trueswell, J. (2009). Use of speaker's gaze and syntax in verb learning. *Language Learning and Development*, **5**, 203–234. 134

Nowak, M.A., Plotkin, J.B. & Krakauer, D.C. (1999). The evolutionary language game. *Journal of Theoretical Biology*, **200**, 147–162. 8

Nowak M.A., J.J.V., Plotkin (2000). The evolution of syntactic communication. *Nature*, **404**, 495–498. 103, 105

Oates, T. (2003). Grounding word meanings in sensor data: Dealing with referential uncertainty. In *Proceedings of the HLTNAACL 2003 workshop on Learning word meaning from nonlinguistic data.*, 62–69, Association for Computational Linguistics. 65

Oliphant, M. & Batali, J. (1997). Learning and the emergence of coordinated communication. *The newsletter of the Center for Research in Language*, **11**. 10

Oudeyer, P.Y. (2005). The self-organization of speech sounds. *Journal of Theoretical Biology*, **233**, 435 – 449. 37

Pantel, P. & Lin, D. (2002). Discovering word senses from text. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '02, 613–619, ACM, New York, NY, USA. 227

Pauw, S. & Hilferty, J. (2012). The emergence of quantifiers. In L. Steels, ed., *Experiments in Cultural Language Evolution*, John Benjamins, Amsterdam. 12

Piccin, T.B. & Waxman, S.R. (2007). Why nouns trunp verbs in word learning: New evidence from children and adults in the human simulation paradigm. *Language learning and development*, **3**, 295–323. 68, 72

Pinker, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge UP, Cambridge. 62, 66

Pinker, S. (1998). Words and rules. *Lingua*, **106**, 219–242. 4

Pollard, C. & Sag, I.A. (1994). *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago. 227

Quine, W. (1960). *Word and Object*. MIT Press, Cambridge, MA. 61

Rakison, D.H. & Oakes, L.M. (2003). *Early category and concept development: Making sense of the blooming buzzing confusion*. Oxford University Press, New York. 135

Röfer, T., Dahm, I., Düffert, U., Hoffmann, J., Jüngel, M., Kallnik, M., Lötzsch, M., Risler, M., Stelzer, M. & Ziegler, J. (2004). Germanteam 2003. In D. Polani, B. Browning & A. Bonarini, eds., *RoboCup 2003: Robot Soccer World Cup VII*, vol. 3020 of *Lecture Notes in Artificial Intelligence*, Springer, Padova, Italy, more detailed in GermanTeam RoboCup 2003. Technical Report (199 pages, http://www.germanteam.org/GT2003.pdf). 180

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, **7**, 573–605. 158

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M. & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, **8**, 382–439. 116, 117

Schober, M.F. (1993). Spatial perspective taking in conversation. *Cognition*, **47**, 1–24. 192

Scott-Philips, T. & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Science*, **14**, 411– 417. 17

Scott-Phillips, T.C. & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, **14**, 411 – 417. 17

Simon, H.A. & Lea, G. (1974). Problem solving and rule induction: A unified view. *Knowledge and Cognition*, 105–128. 65

Siskind, J.M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, **61**, 39–91. 22, 62, 66, 106, 139, 224

Smith, A.D.M. (2001). Establishing communication systems without explicit meaning transmission. In *Advances in Artificial Life Proceedings of the Sixth European Conference, ECAL 2001*, vol. 2159 of *Lecture Notes in Artificial Intelligence*, 381–390, Springer Verlag, Berlin. 180

Smith, A.D.M. (2003a). Intelligent meaning creation in a clumpy world helps communication. *Artificial Life*, **9**, 175–190. 23, 88, 91, 93, 94, 98

Smith, A.D.M. (2003b). Semantic generalisation and the inference of meaning. In *ECAL03*, 499–506. 66

Smith, A.D.M. (2005a). The inferential transmission of language. *Adaptive Behavior*, **13**, 311–324. 62, 66, 106, 180

Smith, A.D.M. (2005b). Mutual exclusivity: communicative success despite conceptual divergence. In M. Tallerman, ed., *Language Origins: perspectives on evolution*, 372–388, Oxford University Press. 99

Smith, A.D.M. (2007). Language change and the inference of meaning. In C. Lyon, C. Nehaniv & A. Cangelosi, eds., *Emergence of Communication and Language*, 323 – 337, Springer. 17

Smith, K., Kirby, S. & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, **9**, 371–386. 17

Smith, K., Smith, A.D.M., Blythe, R.A. & Vogt, P. (2006). Cross-situational learning: a mathematical approach. In P. Vogt, Y. Sugita, E. Tuci & C. Nehaniv, eds., *Symbol Grounding and Beyond: Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication*, 31–44, Springer Berlin/Heidelberg. 22, 66, 83, 88, 98, 106

Smith, K., Smith, A.D.M. & Blythe, R.A. (2009). Reconsidering human cross-situational learning capacities: A revision to yu & smith's (2007) experimental paradigm. In *Proceedings of the 31th Annual Conference of the Cognitive Science Society*, 2711–2716, Cognitive Science Society, Austin, TX. 62, 68

Smith, L. & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, **106**, 1558–1568. 68, 72

Sorensen, T. (1957). A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. *Biologiske Skrifter / Kongelige Danske Videnskabernes Selskab*, **4**, 1–34. 115

Spranger, M. (2008). World models for grounded language games. 180

Spranger, M. (2012). The co-evolution of basic spatial terms and categories. In L. Steels, ed., *Experiments in Cultural Language Evolution*, John Benjamins, Amsterdam. 12, 179

Spranger, M., Pauw, S. & Loetzsch, M. (2010). Open-ended semantics co-evolving with spatial language. In A. Smith, M. Schouwstra, B. de Boer & K. Smith, eds., *The Evolution of Language (EVOLANG 8)*, 297–304, World Scientific, Singapore. 12

Sproat, R. & Shih, C. (1988). Prenominal adjective ordering in english and mandarin. *NELS*, **18**, 465 –489. 199, 202

Sproat, R. & Shih, C. (1991). *The cross-linguistic distribution of adjective ordering restrictions*. Kluwer Academic Publishers, Dordrecht. 202

Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life*, **2**, 319–332. 8, 34, 35, 65

Steels, L. (1996). Self-organizing vocabularies. In C.G. Langton & K. Shimohara, eds., *Proceeding of the Fifth International Workshop on the Synthesis and Simulation of Living Systems*, MIT Press, Cambridge, MA. 34

Steels, L. (1998a). The origins of ontologies and communication conventions in multi-agent systems. *Journal of Agents and Multi-Agent Systems*, **1**, 169–194. 35, 180

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. 16

Steels, L. (1999). *The Talking Heads Experiment. Volume 1. Words and Meanings*. Special pre-edition for LABORATORIUM, Antwerpen. 48, 66

Steels, L. (2000). 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. 4, 48, 128

Steels, L. (2001). Language games for autonomous robots. *IEEE Intelligent Systems*, 16–22. 65

Steels, L. (2002). Grounding symbols through evolutionary language games. In A. Cangelosi & D. Parisi, eds., *Simulating the Evolution of Language*, 211–226, Springer Verlag, London. 8

Steels, L. (2003). Language re-entrance and the inner voice. *Journal of Consciousness Studies*, **10**, 173–185. 10, 111

Steels, L. (2004a). The autotelic principle. In I. Fumiya, R. Pfeifer, L. Steels & K. Kunyoshi, eds., *Embodied Artificial Intelligence*, vol. 3139 of *LNAI*, 231–242, Springer, Berlin. 28

Steels, L. (2004b). Social and cultural learning in the evolution of human communication. In D.K. Oller & U. Griebel, eds., *Evolution of Communication Systems: A Comparative Approach*, 69–90, MIT Press, Cambridge, MA. 3

Steels, L. (2005). What triggers the emergence of grammar? In A. Cangelosi & C. Nehaniv, eds., *AISB'05: Proceedings of EELC'05*, 143–150, AISB, Hatfield. 28

Steels, L., ed. (2011a). *Design Patterns in Fluid Construction Grammar*. John Benjamins. 28

Steels, L. (2011b). Introducing fluid construction grammar. In L. Steels, ed., *Design Patterns in Fluid Construction Grammar*, John Benjamins, Amsterdam. 227

Steels, L., ed. (2012a). *Experiments in Cultural Language Evolution*. John Benjamins. 12

Steels, L. (2012b). Self-organization and selection in language evolution. In L. Steels, ed., *Experiments in Cultural Language Evolution*, John Benjamins, Amsterdam. 12, 13

Steels, L. & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, **28**, 469–489. 12, 21, 50, 65, 221

Steels, L. & Kaplan, F. (1998a). Spontaneous lexicon change. In *Proceedings COLING-ACL 1998*, 1243–1250, Morgan Kaufmann, San Francisco, CA. 34, 35

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. 35

Steels, L. & Kaplan, F. (1999). 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. 48, 66

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. 48

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. 12, 179, 192

Steels, L. & Loetzsch, M. (2012). The grounded naming game. In L. Steels, ed., *Experiments in Cultural Language Evolution*, John Benjamins, Amsterdam. 12, 34, 222, 226

Steels, L. & McIntyre, A. (1999). Spatially distributed naming games. *Advances in Complex Systems*, **1**, 301–323. 21, 33, 35, 54

Steels, L. & van Trijp, R. (2011). How to make construction grammars fluid and robust. In L. Steels, ed., *Design Patterns in Fluid Construction Grammar*, John Benjamins, Amsterdam. 8

Steels, L. & Wellens, P. (2006). How grammar emerges to dampen combinatorial search in parsing. In P. Vogt, ed., *EELC 2: Symbol Grounding and Beyond*, vol. 4211 of *LNAI*, 76–88, Springer Verlag, Berlin Heidelberg. 227

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.

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. 106

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. 227

Sutton, R. & Barto, A. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, **88**, 135 – 170. 51

Tannen, D. (1980). *A Comparative Analysis of Oral Narrative Strategies: Athenian Greek and American English*. Ablex, Norwood, New Jersey. 6

Theisen, C., Oberlander, J. & Kirby, S. (2010). Systematicity and arbitrariness in novel communication systems. *Interaction Studies*, **11**, 14–32. 17

Timberlake, A. (1977). Reanalysis and actualization in syntactic change. In C. Li, ed., *Mechanisms of Syntactic Change*, 141–177, University of Texas Press, Austin. 224

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. 179, 180

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press, Harvard. 140, 141

Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. In M. Bowerman & S.C. Levinson, eds., *Language Acquisition and Conceptual Development*, 132–158, Cambridge University Press, Cambridge. 18, 140, 203

Tomasello, M. (2003). *Constructing a Language. A Usage Based Theory of Language Acquisition*. Harvard University Press. 15, 141

Tomasello, M. & Carpenter, M. (2005). *Intention reading and imitative learning*, 133–148. MIT Press, Cambridge, MA. 134

Traugott, E.C. & Heine, B., eds. (1991). *Approaches to Grammaticalization*, vol. 1 of *Typological Studies in Language 19*. John Benjamins, Amsterdam. 194, 224

Van Looveren, J. (1999). Multiple word naming games. In *Proceedings of the 11th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC '99)*, Maastricht, the Netherlands. 106

Van Looveren, J. (2005). *Design and Performance of Pre-Grammatical Language Games*. Ph.D. thesis, Vrije Universiteit Brussel, Brussels. 222

Van Segbroeck, S., de Jong, S., Now, A., Santos, F.C. & Lenaerts, T. (2010). Learning to coordinate in complex networks. *Adaptive Behavior*, **18**, 416 –427. 36

van Trijp, R. (2008a). *Analogy and Multi-Level Selection in the Formation of a Case Grammar. A Case Study in Fluid Construction Grammar*. Ph.D. thesis, Universiteit Antwerpen. 12

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. 12, 34

van Trijp, R. (2011). The evolution of case systems for marking event structure. In L. Steels, ed., *Experiments in Cultural Language Evolution*, John Benjamins, Amsterdam. 12

van Trijp, R. (2012). Towards culturally acquired language strategies. In M. Hild & L. Steels, eds., *Language Grounding in Robots*, Springer, New York. 8

Vikram K., J. & Markman, E.M. (2001). Learning proper and common names in inferential versus ostensive contexts. *Child Development*, **72**, 768–786. 61

Vogt, P. (2000). *Lexicon Grounding on Mobile Robots*. Ph.D. thesis, Vrije Universiteit Brussel. 48, 65, 66

Vogt, P. (2001). Bootstrapping grounded symbols by minimal autonomous robots. *Evolution of Communication*, **4**, 87–116. 226

Vogt, P. (2002). The physical symbol grounding problem. *Cognitive Systems Research*, **3**, 429–457. 226

Vogt, P. (2003). Anchoring of semiotic symbols. *Robotics and Autonomous Systems*, **43**, 109–120. 226

Vogt, P. (2005a). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence*, **167**, 206 – 242. 17

Vogt, P. (2005b). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence*, **167**, 206–242. 103, 105

Vogt, P. & Coumans, H. (2003). Investigating social interaction strategies for bootstrapping lexicon development. *Journal of Artificial Societies and Social Simulation*, **6**. 21, 23, 48, 51, 68, 71, 88, 91, 93, 98, 221, 222

Vogt, P. & Divina, F. (2007). Social symbol grounding and language evolution. *Interaction Studies*, **8**, 31–52. 68, 106

Vogt, P. & Smith, A.D.M. (2005). learning colour words is slow: a cross-situational learning account. *Behavioral and Brain Sciences*, **28**, 509–510. 66

Wang, S. & Baillargeon, R. (2008). Can infants be "taught" to attend to a new physical variable in an event category? the case of height in covering events. *Cognitive Psychology*, **56**, 284–326. 175

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. 167

Wellens, P. (2011). *Organizing Constructions in Networks*, 181–202. John Benjamins.

Wellens, P. & Beule, J.D. (2010). Priming through constructional dependencies: a case study in fluid construction grammar. In *The Evolution of Language (EVOLANG8)*, 344–351, World Scientific.

Wellens, P. & Loetzsch, M. (2012). Multi-dimensional meanings in lexicon formation. In L. Steels, ed., *Experiments in Cultural Language Evolution*, John Benjamins, Amsterdam. 158

Wellens, P., Loetzsch, M. & Steels, L. (2008). Flexible word meaning in embodied agents. *Connection Science*, **20**, 173–191. 158

Wilkinson, K.M. & Mazzitelli, K. (2003). The effect of missing information on childrens retention of fastmapped labels. *Journal of Child Language*, **30**, 47–73. 61

Wittgenstein, L. (1967). *Philophische Untersuchungen*. Suhrkamp, Frankfurt. 6, 33

Yarowsky, D. (1995). Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, ACL '95, 189–196, Association for Computational Linguistics, Stroudsburg, PA, USA. 227

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, **8**, 338–353. 158

Zempleni, M.Z., Renken, R., Hoeks, J.C., Hoogduin, J.M. & Stowe, L.A. (2007). Semantic ambiguity processing in sentence context: Evidence from event-related fmri. *NeuroImage*, **34**, 1270 – 1279. 9

Zuidema, W. & Westermann, G. (2003). Evolution of an optimal lexicon under constraints from embodiment. *Artificial Life*, **9**, 387–402. 115, 150