Universiteit Utrecht

FACULTY OF THE HUMANITIES
DEPARTMENT OF PHILOSOPHY
COGNITIVE ARTIFICIAL INTELLIGENCE

MASTER'S THESIS

Gesturing Orders and their Interpretation

July 9, 2010

Author:
Anouschka van Leeuwen
3068099

First Supervisor and Reviewer:
Prof. dr. Henriette de Swart

Second Supervisor and Reviewer:
Marieke Schouwstra MSc

Third Reviewer:
dr. Menno Lievers

Abstract

This thesis offers a description of two sets of experiments that were carried out in the field of the evolution of language. Using gesturing as a 'window' to study early language systems, participants are involved in two tasks: 1) gesturing simple events, and 2) interpreting other people's gestures in a multiple choice set-up. The aim is to defy other researcher's claims that only one gesturing order is to be found, and to prove that once multiple gesturing orders are identified, that communication is a significant factor in the way improvised utterances are formed. The results from the studies are then placed in the broader framework of language evolution. To conclude, there is a discussion of hypotheses about the way the semantic and the computational systems of language evolved.

Acknowledgments

Many thanks go to my two supervisors, Henriette and Marieke. When I first approached Henriette with a possible subject for my thesis, she took the liberty of inviting Marieke to our appointment, which led to me starting to work in the field of evolution of language. I owe them both a lot of gratitude: Henriette especially for her enthusiasm, boldness, and trust in the value of the experiments we carried out, and Marieke for her openness and willingness to share her ideas. She was always ready to see me and give me feedback, and when her health did not permit it, we simply continued the discussions at her home. Furthermore, I appreciate the interest they took in me; not only in my thesis, but also in the job interviews I went through, and the opportunity they gave me to be involved in the organization of EvoLang.

The same enthusiasm was found in the third reviewer, Menno Lievers, who did not hesitate to accept my request to be the last reviewer, and who also kept in touch about the situation surrounding my possible future PhD position.

The experiments I carried out would not have been possible without the cooperation of many hard working students visiting the University Library, who generously offered me 5 minutes of their time. My roommates acted as volunteers too, so I would like to thank them as well.

Last of all I like to mention someone who was a big help in matters both related and unrelated to this thesis. He has made available his support in a number of ways: as an actor for the gesturing clips in the first tries at the production experiments, as a technician in designing the applications that were used to carry out the interpretation experiments, and, most importantly, as a friend. Offering me constant support, I want to thank Martijn, who, in line with the subject of this thesis, tells me the most when he does not use any words at all.

Contents

Part I

Foundations

Chapter 1

General Framework

1.1 CAI and the Importance of Language

This thesis is written for the master program Artificial Intelligence (AI), a multidisciplinary study that combines psychology, computer science, logic, linguistics, and epistemology. The goal is to study intelligence and apply that knowledge in technical applications. An interplay exists between human and computer, because both can be used to study each other. What aspects of human intelligence can be applied in computers, and how can we use computer models to gain more knowledge about human intelligence? Those are the core questions of AI.

An important part of human intelligence is the ability to gain, and reason with, knowledge. Knowledge, in turn, is heavily dependent on language: it is the 'carrier' of knowledge, and also has a intricate relation with thought and reasoning. Instead of studying one particular language, theoretical linguistics deals with language *in general*.

The area of linguistics that is the context of this thesis, is the study of the evolution of language. Like AI, it is a very broad and interdisciplinary field. It has applications in almost any of the subfields of AI, which makes it such an elementary part of it. For example, computer models can be used to study evolutionary processes, and also, psychology and biology play a huge role in language evolution. Modern theories about language development will inevitably contain models about the cognitive structure of our minds, which is - again - an important issue in AI.

In the next section, a short overview will be given of the study of the evolution of language. In particular, the aspects that are of importance to the research questions of this thesis (see section 1.3.1) will be discussed.

1.2 Evolution of Language

1.2.1 Short Overview

Countless overviews of the language of evolution have already been written. This particular one is an overview in the sense that it gives a description of the questions that are of importance to the subject of this thesis. This is necessary in order to be able to compare the experiments that were carried out for this thesis, to work by other researchers. There are some differences in the approaches that are used, which makes it useful to examine the theoretical basis beforehand. The works by Kirby ([14]) and Johansson ([13]) are used as a guideline here, because in their overview they pay attention to gesturing, a relatively new research area.

There is consensus that no matter how much of language is learned, it is clear that language is enabled (and constrained) by human biology. This brings us to the first question: what does the system of language in our body look like? What parts does it consist of? Language not only involves the structure of sentences, but also the meaning of individual words, the uttering of speech, etc. Should this complex system be treated as a unitary whole, or as separate systems?

Hauser, Chomsky & Fitch ([10]) divide language in two faculties: the narrow one, which is essential to language and involves the computational system, and the broad one, which contains the remaining systems like the organs responsible for speaking. They hold that the narrow faculty is all that is uniquely human. Other authors are not quite as strict in their division. In general, there seems to be a consensus that three systems are particularly important to the study of language: the (just mentioned) computational system, the conceptual system, and the perceptual-articulatory apparatus. However, even within this "common" view, a lot of variations exist about the relative importance of each system and the strictness of division between the three.

Thus, one part of the debate is about the exact nature of language. The second question is how language came to be. Apart from the question whether separate modules exist, the question is whether there was a gradual evolution of the language faculty, or whether there were 'revolutions' that made language possible as we know it today. If we take the modules-view, the question becomes whether there was a particular order in which the modules developed, for example, did the computational system evolve at an earlier stage than the conceptual one did? Intermediate stages in the evolution of language are a popular idea, i.e. 'protolanguages' are often spoken of.

Furthermore, another question is *why* language evolved. Had the process of evolving a language system an adaptive function, so that natural selection could come in? Or was language simply a byproduct (a spandrel, as it is called by biologists)? Chomsky ([3]) views the existence of the computational system as a spandrel, which gave rise to human language abilities. He finds a more gradual, adaptive view unlikely: "In the case of such systems as language or wings it is not easy even to imagine a course of selection that might have given rise to them." Other theories do favor language as an adaption (Pinker & Bloom, [17]). But, as Kirby notes, an adaptationist view receives a lot of skepticism. After all, intermediate stages of a complex system such as language do not make sense unless each of them has a function of its own, just like a half-developed eye is useless: "A gradual story for the evolution of the eye is only possible if it turns out that there are 'intermediate' eyes that are indeed useful and there is a plausible evolutionary trajectory from these intermediate forms to the modern eye." Thus, if a gradual evolution of language is proposed, it should be backed up with functional intermediate stages. There is no consensus about what the driving force behind the gradual evolution is; possibilities are social grooming, sexual display, communicative purposes, and alliance-forming. Also, the possibilities for the intermediate stages are not necessarily *lingual* in nature, but could also involve a stage in which gestural movements were used (which will be extensively discussed below).

To come to the last important issue, as was said there is a lot of variation between languages. However, there are also quite some commonalities, and some limits on what linguistic structures are possible within language. The question is how much of the structure of language is genetically determined, and how much is learned. There are some arguments that would point to language being largely innate, the most well known of which is 'the poverty of the stimulus'. As Johansson describes it: "as the space of all possible grammars is infinite, it is impossible in principle to identify the target grammar of acquisition without innate constraints on the search space".

To summarize, important questions regarding the evolution of language include the following contrasts: modules-view versus unitary whole, gradual versus sudden evolution, spandrel versus adaptive function, innate versus learned language, and the question whether there was a non-linguistic predecessor. For most of these contrast, the existing theories are a combination of both sides. One idea that is accepted by most researchers is that language evolution was gradual. This would mean that there were a number of intermediate stages, or protolanguages, during the evolution of language. As Johansson notes, it is of utmost importance to keep in mind that each step should be an advancement on its own - and we should not view evolution as ascending of stairs with a common goal. Evolution only takes place if a change has an adaptive function.

Johansson in his book sets out the possible combinations of the views expressed above. His conclusion is that the most promising theories off language evolution are the ones that support a gradual, adaptationist view, in which language developed early in human history.

1.2.2 The Windows Approach

The questions surrounding language evolution are all very important, but some of them are hard, if not impossible, to study. How can the first occurrences of languages be described, when obviously, those situations are not available for observation any more? Claims about the emergence of language in a distant past have no direct evidence to back it up, which is a big obstacle (also called the 'data problem').

Thus, researchers have to find a way around this handicap. Indirect data that might be include computational models that may give insight into the processes of evolution, or the observation of animals and animal communication, who after all share some of our genes. A more recent idea is to study 'living fossils': types of communication used by modern humans that are close to, but do not share all the features of, fully-modern language ([1]). The assumption is that since there is no direct evidence to work with, *in*direct evidence could be used instead. Phenomena that are related to the evolution of language, and still observable today, can be used to gain insight into language evolution.

Botha ([2]) has crystallized this idea into a more formal method called the 'Windows Approach'. Features of language evolution can be 'seen' by 'looking at' it through 'windows' offered by various phenomena. Windows not only include objects like fossil skulls and ancestral brains, but also 'living' systems like modern motherese, pidgin languages, homesign systems and second-language learners.

Botha translates the methaporical language about windows as follows: "to - metaphorically - 'see' a property of some aspect of language evolution by - metaphorically - 'looking at' a property of some other phenomenon is - non-metaphorically - to infer the first property from data or assumptions about the second property". For example, by studying pidgin languages we wish to make inferences about possible stages of protolanguages.

Not every phenomenon related to language counts as a window. Botha has formulated a number of conditions windows should adhere to. First of all, Groundedness, which means the window needs to be a well understood phenomenon, including having an underpinning theory and descriptions of the assumptions from which inferences are drawn. Next to that, there is Pertinence, which means windows should be about the 'right kind' of evolution of language. For example, it should be about the emergence of early language, not about changes in fully fledged modern languages. A third condition is Warrantedness, meaning that the inferential steps leading to some conclusion about what language evolution involved need to be suitably licensed. This can be done by using bridge theories, in which phenomena in two distinct ontological domains are interrelated in a way that makes it possible to move inferentially from the one to the other. The conditions are quite demanding, and therefore a lot of windows that are presently being investigated do not yet 'officially' count as a proper window. Another complicating factor is that windows may differ regarding the aspect of language evolution about which conclusions are drawn, and the inferential steps by which those conclusions are drawn. Conclusions about pidgins, for example, are the results of an analogous inference, because they are seen as an analogue for protolanguage. Another inference may draw on correlations, if for instance two phenomena precede each other in time or if there is causation between them.

To conclude this section about the windows approach, it is clear that this area of research is still relatively young, and needs a lot of refining. It is also clear that there is a lot of heuristic potential here, because it enables researchers to study situations that were thought to be no longer available. It is a promising way to study the evolution of language.

1.2.3 Gesturing

The experiments that were carried out in the context of this thesis, made use of a window on language evolution as well. That window is *gesturing*: the act of moving the limbs or body as an expression of thought. As was noted in the previous section, some windows are still quite young and therefore the exact description that are needed to confirm to Groundedness, Pertinence, and Warrantedness, may be lacking. This is the case as well for the gesturing window. In fact, gesturing is not even fully recognized yet as being a window at all. More often, gesturing is studied not because of its potential for being a window on language evolution, but because it as seen as a potential predecessor of language.

Gesturing as an intermediate stage

Remember from the discussions in the previous sections that a gradual view of language evolution implies the existence of intermediate stages, or protolanguages, which are the ancestors of modern language. Taking this to the very beginning, was the first stage a protolanguage that consisted of spoken language, or did the first stage involve another modality (another type of sensation, such as feeling or hearing)? As Johansson describes, there are a few reasons why gesturing could be the first stage from which spoken language evolved later on. The main reason is that fully developed sign languages share features with spoken languages: they are both advanced grammatical systems, and share the same stages during development (from babbling, to basic vocabulary, etc.). The researcher Susan Goldin Meadow has done a lot of experiments in the field of sign languages. She studied the emergence of pidgin languages among deaf people, which showed fundamental combinatorial properties also seen in spoken language. Furthermore, gesturing is used by 'normal' speakers as well. It is almost always present during speaking to strengthen spoken messages, even when talking to blind people. So, it seems that gesturing has a tight bond with speaking.

In the context of language evolution, our closest relatives, primates, are also capable of gesturing, which makes an evolutionary tract starting from gestures in apes towards spoken language in humans a plausible option. The question is: why and when did we switch to speech? It could also be that gesturing and speaking were used alongside each other for a while, eventually developing into a system where speaking became dominant. But, like the other general questions, none of these speculations have been proven yet.

Gesturing as a window

Gesturing was used in the windows context for the first time when Goldin Meadow and her colleagues published an article describing a new experiment ([8]). In this experiment, instead of studying deaf people and their gesturing systems, *speaking* participants were recruited to carry out a gesturing task. They had to watch a picture of a simple situation, and then describe it using gestures only. The result was that speakers almost exclusively used the ordering Actor-Patient-Act, which corresponds to Subject-Object-Verb, for example boy-glass-drinks. The results were surprising because the participants came from two groups: one consisting of speakers whose mother language has predominant order Subject-Object-Verb, and the other Subject-Verb-Object, so the mother language's grammar did not influence their gesturing.

Goldin Meadow et al. want to use the results to draw conclusions about the relation between ordering rules of modern language and nonverbal representations of events, so to study the cognitive basis of the representation of events. In the last paragraph, they suggest that the results may also be important to draw conclusions about situations in which new languages are created, but overall, their focus in not on language evolution.

However, from their article the insight has been derived that gesturing could very well be a window as we described them within the windows approach. Since speakers bypassed their native speaking order, it means that a gesturing task simulates a situation in which no grammar is used, and instead, simple improvised communication is required.

Interestingly, Goldin Meadow does not term the gesturing task a communicative one, but 'merely' labels it non-verbal. But the communicative aspect of improvised gesturing is exactly what will become important later on in this thesis. We want to use gesturing as an analogous window: by letting people gesture, we create a pre-linguistic, grammar-less condition. Gesturing has some parallels with other windows like pidgins, homesigning systems, and second language learners. All of these situations 1) demand a great deal of improvisation, 2) contain a very restricted set of linguistic elements (be it words or gestures), and 3) there is no possibility to rely on your mother language.

What will people do when they are forced to communicate with no advanced system existing yet? This brings us to the research questions of this thesis.

1.3 Thesis

1.3.1 Research Questions

The work by Goldin Meadow and others will be discussed more extensively in the next chapter. For now, it is enough to explain that the stimuli that Goldin Meadow used in the gesturing task - the pictures that the participants had to gesture - are uniform: they all contained the same kind of verb. The kind of pictures she used are not the only kind of simple situation that one could think of. Thus, it is worth our while to investigate what will happen when other stimuli are presented to participants. This leads to the first research question.

Research Question 1: Will stimuli that differ semantically from the ones used by Goldin Meadow et al. lead to different gesturing behavior in normal hearing and speaking participants?

As will be explained in the next chapter, there already exists some evidence that the result from Goldin Meadow is not the complete story, and that when the stimuli are semantically modified, the gesturing that results will be different too. If we obtain this result again after examining the first research question, a new issue comes up. We claim that once you find several gesturing orders - not just one- communication will inevitably play a role. To confirm this hypothesis, the following question is formulated.

Research Question 2: Do the gesturing orders (the answer to question 1) have a communicative function, so that they will systematically lead to different interpretations when presented to normal hearing and speaking participants?

Question 1 will be answered by carrying out so called ’production’ experiments, in which the participants are asked to gesture. Question 2, in turn, reverses that set-up, and lets people watch other people gesture, which leads to ’interpretation’ experiments.

1.3.2 Outline

This thesis has the following structure. Chapter 2 gives a detailed discussion of the research that has been carried out using gesturing as an experimental task. Also, the research questions posted above will be described in greater detail. Then in Part II, the experiments that were carried out are described, separated into two chapters: one about Production, and the other about Interpretation. At the end, Part III contains an overall discussion and conclusion of the results. Furthermore, there is a section about further research based on the current work.

Chapter 2

Discussion of Literature

2.1 The case for SOV

In section 1.2.3, the study by Goldin Meadow et al. was already mentioned. Their experiment was duplicated in part by Langus & Nespor ([15]). In the sections below, their methods are separately described, because the paradigm that these authors work in differ from each other, as do their conclusions. With respect to the introduction to the evolution of language, these authors take a different stance on several issues discussed earlier.

2.1.1 Goldin Meadow

The work by Goldin Meadow is seated in the study about sign languages, particularly about the way children learn such a language. She is interested in what happens when people learn a new (sign)language without any external linguistic influences, for example, deaf children who are not displayed to a conventional sign language.

In their groundbreaking article from 2008 about gesturing, Goldin Meadow et al. do not explicitly state the relation of their work to the study of the evolution of language. As was said before, the recognition that gesturing may be used as a window is still relatively new. There is also no mention of separate modules: the aim of the experiment is to investigate the relation between ordering rules of modern language and nonverbal representations of events, without stating that their experiment will delve into a particular language faculty.

The experiment was carried out as follows: there were two groups of participants who differed in the dominant word order of their mother language. One group were speakers of Subject-Object-Verb (SOV) languages, the other from Subject-Verb-Object (SVO) languages. They were asked to carry out a nonverbal task: after being shown a picture of a simple event (for example, 'a woman twists a knob'), they had to describe the picture without speaking, i.e., by gesturing.

The result showed that both groups used subject-object-verb order. To make a difference between cognitive structures and uttering structures (gestures or words), Goldin Meadow uses the terms Actor (for subject), Patient (for object), and Act (for verb) for the cognitive realm, and S, V, and O for utterances. The result is surprising, because the groups differed in their mother language. Thus, there seems to be a universal way to gesture simple events.

The conclusions that Goldin Meadow et al. draw from their experiment are the following. First of all, that the ordering found in a speaker's habitual talk does not inevitably influence that speaker's nonverbal behavior. Furthermore, from Goldin Meadow's work in sign languages, she links the finding of SOV to the finding that deaf children often start out with SOV as well, thus suggesting that Actor-Patient-Act is a natural order for representing events, and one that is used when creating a new (sign) language. She thus does not suggest that SOV is used for communicative purposes. On the contrary, Goldin Meadow is one of the researchers who has a strict view *against* communication being responsible for the SOV order she found. This is already apparent in the way she terms the gesturing task: not as a communicative task, but simply as a nonverbal one.

To back up the claim that communication is not responsible for gesturing orders, a second small experiment was carried out by Goldin Meadow et al. This time, participants again saw the vignettes, but instead of gesturing them, they received three transparent pictures which they had to lay on top of each other to form the original vignette. The transparent pictures showed the subject, the object, and the verb. The result was that again, participants used SOV, so they first stacked the subject, then the object, and then the verb. Goldin Meadow et al. now reason that since this task is non-communicative, SOV can not have a communicative function. We will come back to this second task below.

2.1.2 Langus & Nespor

Two years later, Langus and Nespor carried out a set of four experiments, of which the first is a duplication of the experiment by Goldin meadow et al., with the same result: a clear SOV preference when people gesture events. Their work is seated in a different framework, heavily influenced by the modules-view of language, and by the work of Chomsky and others ([10]). In their article, the division between the conceptual and the computational system is of primary importance. The tasks of those systems are strictly determined: the conceptual system deals with semantics and interpretation, while the computational system deals with grammar: "We rely on the proposal that the human faculty of language is modular and that it is possible to identify different cognitive systems responsible for specific linguistic tasks". Those tasks are associated with different structures of representations. In line with Goldin Meadow, the conceptual system receives Actor-Patient-Act (ArPA), while the computational one uses Actor-Act-Patient (ArAP). One of the examples that Langus and Nespor use is the difference between the development of pidgin languages, which are spoken, "fully fledged" grammatical languages, and the development of a sign language, which was referred to earlier. The first shows a SVO development, while the second has a SOV structure, which shows the "dichotomy" of the two structures.

In their article, Langus and Nespor are fully aware of the insights that gesturing tasks can offer. While Goldin Meadow was still reluctant to relate her work to the evolution of language, Langus and Nespor clearly state that they view gesturing as a prelinguistic task. They emphasize the absence of grammar during gesturing. During this task, "the mapping between the signal (the gestures) and its meaning is achieved without the intervening syntactic computations responsible for phrase structure". In other words, they state that they have found a direct interaction between the conceptual system and the articulatory one, without grammar playing a role. Their duplication of the gesturing experiment confirms their thoughts, because again SOV is found as the dominant order. Furthermore, they differ from Goldin Meadow because they see gesturing as a pre-linguistic *communicative* task, while Goldin Meadow termed it a nonverbal task, in which communication is not mentioned.

A new element that is introduced to the gesturing research is that not only a *production* experiment is conducted (i.e. letting people gesture events), but also one about *interpretation*. Instead of letting participants describe events, they had to watch someone else who was gesturing. Again, the participants came from languages that differed in predominant word order. After watching a short movie clip of someone gesturing, two pictures appeared that the participants could choose from, and they had to pick the one that had just been gestured in the movie. The two answers they could choose from differed randomly in either the subject, the object, or the verb. The gesturing clips were recorded in different gesturing orders, among which SOV and SVO. The goal was to determine which gesturing order would lead to the fastest correct responses, i.e. the fastest interpretation.

The results were that SOV led to the fastest comprehension times for all participants. Again, people from different mother languages showed a preference for SOV, even though that is not the predominant order of their language. Langus and Nespor use this study as another confirmation of their earlier claim: participants are not using the computational system, and thus the conceptual system decides, favouring ArPA, thus SOV.

The remaining part of Langus and Nespor's work deals with the claim that the computational system favours ArAP, thus completing their general view that there is a great difference between the conceptual and the computational systems in terms of the representational structures that they contain. This time, another interpretation experiment was conducted, in which not gesturing clips, but prosodically flat sequences of words were used as stimuli. The participants heard a sequence of three words (subject, object, and verb) in all six possible arrangements, and had to choose between two vignettes which were possible representations of the sentence they just heard. The result was that the participants responded most quickly when hearing the order of their native language. This is in contrast with the universal 'computational preference' for ArAP, because the SOV native speakers responded quicker to SOV instead of SVO. Langus and Nespor deal with this problem by comparing the *average* reaction time for orders in which the object precedes the verb (SOV, OVS, OSV) and the orders in which it does not (SVO, VSO, VOS). They then find that the latter group has a lower mean reaction time, thus confirming their hypothesis that ArAP (SVO) is in general best suited for the computational system.

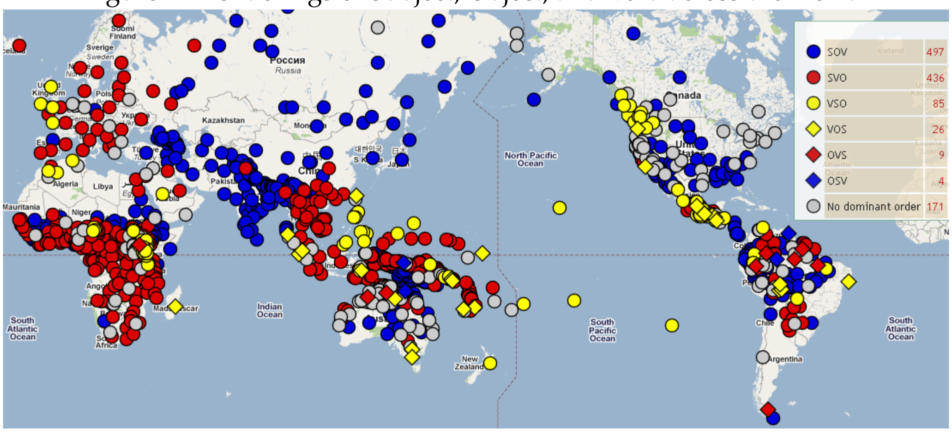
It may be said that this result is not quite convincing. An alternative explanation is that the group of (SOV, OVS, OSV) contains two instances of the subject coming *after* the object (OVS and OSV) while the group of (SVO, VSO, VOS) has only one in which this happens (VOS). Following Greenberg ([9]), linguistic orders in which the object precedes the subject are very rare and thus harder to interpret. This may account for the fact that the group in which the object came after the verb had an advantage over the other, because that was also the group in which the subject preceded the object twice.

Langus and Nespor do not discuss this possibility and stick to their theory: the conceptual domain favours ArPA (SOV), and the computational system favours ArAP (SVO).

2.1.3 SOV versus SVO

The central structures in the work by Langus and Nespor are SOV and SVO. It is interesting to see the 'state' of these orders today: in figure 2.1, a map is shown of the dominant word orders of languages across the world ([4]).

Figure 2.1: Orderings of Subject, Object, and Verb across the world.



SOV and SVO occur most frequently, and there are also quite a few languages that do not have a dominant order. Langus and Nespor make a very strict distinction between the conceptual system and grammar, and hold that the latter favours SVO, which accounts for the existence of SVO languages. However, SOV is tied to the conceptual system, so why would it be present in the grammar of so many languages? As an explanation for their 'dilemma', Langus and Nespor think that conceptual forces somehow interfered with grammar, which led to the occurrence of SOV languages as well. Their apparent surprise at the number of SOV languages is a result of their view of language evolution. In line with Chomsky, they think that language underwent a sudden development: at one particular point, something in the evolution of language happened that facilitated the existence of grammar. A different view will be defended here, namely that language evolved gradually. But before that view is developed, work is being discussed that sheds new light on the SOV and SVO orderings.

2.2 Semantic pressures

After having discussed the main ideas of Goldin Meadow and Langus & Nespor, it is time to turn to the set-up of their experiments. In particular, what stimuli did they use? These vignettes, or pictures, that the participants were asked to gesture are very important, because those are the ones that tell us about the gesturing orders that resulted.

In both studies, the vignettes displayed simple situations. More exact, the vignettes displayed *motion events*: a situation that in some way involves a movement. This movement is carried out by the subject while staying in place. Examples from Goldin Meadow include 'twisting', 'covering', and 'picking up'. Examples from Langus and Nespor include 'throwing', 'shooting', and 'feeding'.

In explaining the particular events that were used, it becomes clear immediately that *motion* events are not the only kind of event that exist. This is exactly the starting point that motivated Schouwstra ([18]) to investigate what would happen if the production gesturing task was repeated with stimuli that are *not* motion events.

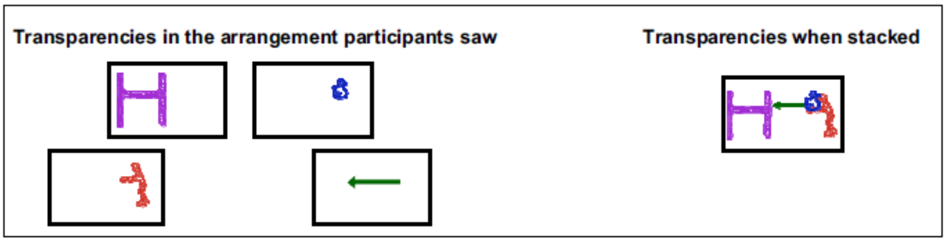
2.2.1 Work by Schouwstra

The simple events that both Goldin Meadow and Langus and Nespor used, consisted of a subject, an object, and a motion (a verb). Within these three constituents, there are a number of semantic properties that may vary, and which may influence the kind of event that they make up together. The particular property of simple events that Schouwstra wanted to examine, is the *kind of verb* that is used. Instead of using motion events, what would happen if *intensional* events were used?

Intuitively, the difference between an intensional verb and a motion verb is the way the object is dependent on the verb. For motion events, both the subject and the object refer to concrete entities that are 'present' in the here and now, so the object that is referred to exists on its own and is not dependent on the verb. Intensional events, on the other hand, involve a dependency of the referred object on the verb: the verb causes the state of the object. For example, in a situation where a subject is 'painting' something, the state of the object is not yet fully determined, and may not even exist in reality. Other examples of intensional verbs are 'dreaming', 'thinking', 'seeing', and 'drawing'.

Remember that the in the second experiment by Goldin Meadow, the participants had to take transparent pictures and lay them on top of each other to form a complete vignette. These pictures did not involve a movement *in* place, but *across* space, which was depicted with an arrow. So all the vignettes consist of objects or persons that relate to each other in a certain way, and this relation is depicted with an arrow (for an example, see figure 2.2). All these items have a very particular informational structure. To see this structure, consider the example "A man moves to a motorcycle". Here, two items (man and motorcycle) are present, and they are related to each other by a 'move-to'-relation. For this example, it feels straightforward, when having to reproduce this event in some order of transparent pictures, to first take the objects, and then the arrow. But what if we take, for example, the following sentence: "A man thinks about a motorcycle". The relationship between the two items in this sentence is not parallel to the 'moves-to' relation. The motorcycle does not exists independently of the man; it is dependent on the man thinking of it. In this example two items are present, but the informational status of the second item depends on that of the first. In this case it feels less straightforward to take both objects first and then the relation between the objects. Similarly, when gesturing such events, it seems more natural to first gesture the verb, and then the object, which is dependent on it.

Figure 2.2: Example of the second experiment by Goldin Meadow et al.



Forbes ([6]) formulates three formal ways of distinguishing an intensional verb, one of which is the ability of substituting one expression for another that is coreferential with it in the complement of the verb, so that the truth-value of the sentence in which the verb occurs can change. For example, Superman is the same person as Clark Kent, but if someone is *painting* Superman (an intensional event), this does not mean that this person is painting Clark. On the other hand, if someone *pushes* Superman (a motion event), this *does* mean that this person pushes Clark.

Schouwstra then set up a pilot study in which people were asked to gesture both motion and intensional events. The stimuli were presented in groups, so that the participants first gestured some motion events, and then some intensional ones. The result was very surprising: for the motion events, SOV clearly dominated, while the intensional events were gestured almost solely using *SVO*. This was clearly a sign that changing the semantics of an event can have an effect on the way it is gestured.

2.2.2 Work by Meir et al.

Within a motion event consisting of subject, object, and verb, the verb is not the only constituent that can be semantically modified. While Schouwstra modified the verb, there are also the subject and the object which possess semantic properties. In a forthcoming study by Meir, Lifshitz, Ilkabasaran, and Padden ([16]), the semantics of the picture of the object in the vignettes were varied. They constructed the stimuli in such a way that there were two kind of clauses, called canonical and reversible clauses. The first kind contains an animate subject and an inanimate object, while the second kind contains two animate arguments.

When such clauses are uttered by using SOV, this will possibly lead to confusion when both arguments are animate. For example, compare "the girl pulled the cart" to "the girl pulled the man". In the first event, only the girl can do the pulling, while in the second, both the girl and the man would be able to do the pulling. Thus, in clauses with only one animate argument, only the subject can actually function as the subject, while two animate arguments lead to ambiguous representations.

Participants were again asked to gesture events, this time a combination of canonical and reversible ones. The result again showed a different picture from the one painted by Goldin Meadow and Nespor and Langus: in clauses with an inanimate object, SOV order was dominant (65% of responses) and SVO appeared only in 31% of the clauses. In clauses with an animate object the reverse pattern was found: the dominant order was SVO (64%), and SOV occurred in 31% of the responses.

This is another piece of evidence that is a contradiction of the claim that ArPA (SOV) is the universal structure for representing events.

2.2.3 Thematic Roles

In the section about Goldin Meadow, the difference between using Ar, A, P, and S, O, V was first mentioned. The first set is frequently used to denote *thematic roles or relations*, which express the semantic relations that the entities within a structure bear towards each other (Harley, 2007). We will continue using this difference to distinguish between cognitive structure and utterance structure.

Thematic roles can be used to clarify the contributions of Schouwstra and Meir et al. If we first look at 'classical' motion events, the thematic relation between Act and Patient is about the trajectory/movement of the Patient across, or in, space. The existence of the Patient is known before the Act is carried out and connects the Actor to the Patient. The difference with intensional events is that in the latter, the thematic relation between Act and Patient is not about movement, but about inherent properties of the Patient. For example, the intensional event 'knitting' is about whether or not the Patient exists (because the scarf is not finished yet), and the intensional event 'thinking' is about the ontological state of the Patient (because the Patient could be an imaginary object).

The work of Meir et al. concerns the possible confusion of thematic roles. They choose the two arguments in such a way that both are suitable to be the Actor and the Patient within the structure. This leads to confusion, which the one who has to gesture apparently solves by using SVO instead of SOV, thus clearly separating the subject and the object by placing the verb in between. Table 2.1 summarizes the contributions of Schouwstra and Meir et al. In the next section we turn to the explanations that the various researchers give for their results.

Table 2.1: Contributions of Schouwstra and Meir et al.

Author(s)	**Introduced modification**
Schouwstra	Variations of Act (A) and relation Act-Patient (A-P)
Meir et al.	Arguments that suit both the Actor and Patient role

2.3 The function of gesturing orders

2.3.1 Cognition versus Communication

The new findings regarding gesturing not only give rise to the question what exactly the basis structure for representing events is, but also to the question what the *function* of all the different orders is. Meir et al. hypothesize that while gesturing orders are indeed cognitive structures, communication is the explanation for the variety between them: "Our study shows that different types of clauses pose different communicative challenges, and different word orders and other devices may emerge to cope with them".

However, not everyone shares this view. Indeed, in his overview of the debate surrounding the evolution of language, Kirby states that the function of communication in this area is still being debated. As was discussed earlier, Goldin Meadow is one of the researchers who has a strict view *against* communication being responsible for the SOV order she found. This was apparent in the way she terms the gesturing task: not as a communicative task, but simply as a nonverbal one. She carried out a second experiment in which a non-communicative task was carried out. Goldin Meadow et al. found the same result as in the gesturing task, SOV, and argues that this proves that communication does not play a role. However, one could argue that this is rather weak evidence: another plausible conclusion is that ArPA is a cognitive structure which is used for communication, but also shows itself in a non-communicative task.

On the other side, Langus and Nespor do mark gesturing as being a communicative task, although they agree with Goldin Meadow et al. that ArPA is a cognitive structure that is not influenced by communication. Their theory is that at some point, the conceptual system along with its ArPA structure came in contact with the computational system, which account for the prevalence of SOV and SVO today.

Table 2.2 shows the positions of the four parties discussed so far (including the author, whose opinion is expressed more clearly in the next section).

Table 2.2: Views on the role communication.

Author(s)	Description of gesturing	Explanation of gesturing orders
Goldin Meadow et al.	Non-communicative	Cognitive only
Langus & Nespor	Communicative	Cognitive only
van Leeuwen	Communicative	Communication, possibly cognitive
Meir et al.	Communicative	Cognitive & influenced by communication

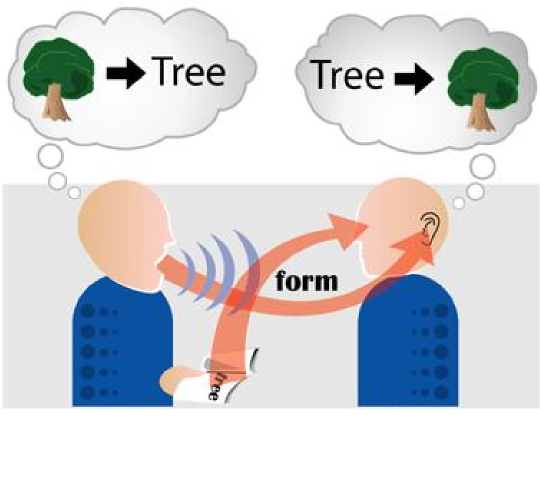
2.3.2 From one to multiple structures

A fundamental point is that Goldin Meadow and Langus and Nespor presuppose that SOV/ArPA is a *unique* structure. As long as only one structure is assumed, communication need not be involved. After all, in that case you can stick to the claim that gesturing is simply a conceptual process in which the only existing conceptual structure for representing events is used for utterances: SOV. The same can then be said of the interpretation of gestures: this process is also a matter of simply mapping the gestures onto the only existing structure: ArPA.

Starting from this claim of uniqueness, the production and interpretation experiments done by Langus and Nespor need not be tied to each other. They can be used separately to back up their claims, as they say themselves: "The preference for the SOV order should not only prevail in gesture production, but also be observable in gesture comprehension".

However, once *multiple* structures emerge, the role of communication can no longer be ignored, which becomes clear when we consider what we take communication to be in more detail. Communication is seen as a process of translating one's thoughts into a particular structure, which is send to the hearer, who then translates that structure into a thought again. The intention of the speaker is to choose the structure of the utterance in such a way that the translation of the hearer will result in the thought the speaker originally had. Figure 2.3 ([20]) shows this process: the picture of the tree is what the speaker wants to convey. To do this, he chooses the uttering 'tree', in the hope that the hearer will translate 'tree' to the mental representation of tree again.

Figure 2.3: Schema of communication.



Supposing that SOV/ArPA indeed is the unique structure for gesturing events, communication will not play a role during gesturing, since the speaker (the one who is gesturing) does not have to keep in mind the translation process of the hearer (the one interpreting the gestures). After all, there is only one way of conveying information, namely SOV.

The studies by Schouwstra and Meir et al. show that it seems to be the case that SOV is not unique in representing events after all. If it can be proven that multiple structures exist, communication will have to be taken into account. Because now, the speaker (the one gesturing) has to contemplate what the hearer (the interpreter) will do with the utterances he receives. The other way around, the interpreter will think about why the speaker chose a particular structure. Relating to the picture, instead of only being able to use 'tree', other forms are available, and the speaker will have to think about which one to use. What Meir et al. would say, is that when there is doubt about who is the subject and who is the object, the speaker chooses SVO to make sure the hearer understands the second argument is the object, and not the other way around.

The way Goldin Meadow chose to use Ar (actor), P (patient), and A (act) instead of S, O, and V, is actually quite fitting here. It shows that the structure of the thought of the speaker (the one gesturing) is not necessarily the same as the structure of the utterance, and that the utterance of the speaker can result in a different thought for the hearer (the interpreter).

To summarize, a single gesturing structure on its own has no 'meaning', but once you have multiple ones, they become significant in the context of communication. A structure can then no longer be studied at only the cognitive level, but has to be considered within the system of cognition *and* communication.

2.3.3 Cognitive structure versus Utterance structure

An important implication of the discussion above is that Langus and Nespor, and Goldin Meadow, draw their conclusions at a different level than is done in this thesis. They suppose that ArPA is a structure *in the brain* for representing events, which is then used as the structure for the gesturing (SOV). They suppose that ArPA is the structure for the thought preceding the gestures. However, in this thesis we will refrain from making statements about the exact structure of mental representations. There is a difference between a complex mental representation, and a linear linguistic expression. The latter has a flat linearized structure, while the first does not. Thus, it is inappropriate to speak of the 'order' within mental representations. The only thing that is claimed, and which we can check, is that the structure of *the utterance* will vary according to the semantic properties of the thought (which are described using thematic roles). What that thought looks like exactly, is unknown.

The grounding of the experiments is therefore fundamentally different. Goldin Meadow and Langus and Nespor wish to investigate the cognitive basis of representing events, and use gesturing to do this, because then grammar will not interfere. They assume that the order of gesturing will originate from the cognitive structure. Gesturing, to them, is a tool.

On the other hand, because we feel that multiple gesturing orders shift the focus towards communication, gesturing orders instead of cognitive orders have become the starting point. On the level of cognitive structure, we remain in the dark, and instead the *function* of gesturing orders becomes the new research question. As in figure 2.3, the mental representation does not have a structure, but is seen as a complete whole, a tree instead of 'tree', which has particular semantic properties. Indeed, the stimuli that participants received in the work of both Goldin Meadow and Langus and Nespor consisted of *pictures*, so the message they had to gesture was originally a 'complete whole' as well.

Throughout this thesis, there is thus a distinction between: 1) the utterance structure, denoted by S, O, V, and 2) the cognitive structure, denoted by Ar, A, P. Goldin Meadow and Langus and Nespor use Ar, A, and P to denote *linear* structures in the brain, while we refrain from such 'linearizations' and instead remain of the opinion that cognitive representations have a complexer built. We will use thematic roles to denote semantic properties of mental representations.

2.4 Aim of the Experiments

2.4.1 Production

In the previous sections, two semantic 'forces' within an event have been described: the kind of verb that is used, and the (in)animacy of the object. Schouwstra already conducted a pilot study in which intensional items were used. It is now time to turn to the experimental part of this thesis. In the first set of experiments, the aim is to examine the gesturing behavior of participants when faced with intensional events in closer detail. It is in fact a duplication of the pilot study by Schouwstra, although some improvements have been implemented. For example, while Schouwstra let participants hear a spoken sentence which they then had to gesture, in the current study the participants are shown vignettes, thus eliminating the possible influence of word order.

The question that will be answered is: do intensional events lead to a different gesturing order than motion events? Relating this to the discussion about communication, the part that will be examined is the way the speaker (the one gesturing) chooses a structure to convey his message. His message, in this case, is manipulated to contain a semantically different kind of verb than the one used in previous studies. On the basis of Schouwstra's work, the hypothesis is that a different structure than SOV will be found.

2.4.2 Interpretation

The production experiments are about the first step of communication: the speaker's message. The second set of experiments is about interpretation, the other end of communication. The experiments evolved from a pilot study to a more formal set-up, the last of which resembles the interpretation experiment done by Langus and Nespor, albeit with some significant modifications.

The question that will be answered is: do the gesturing orders that were found for motion and intensional event result in two different interpretations? Again relating this to communication, in these experiments we move from the utterance of the one who gestures towards the translation process of the interpreter. We investigate whether the structure the speaker chooses, influences the translation of the hearer.

The two possible outcomes of the interpretation experiments are that 1) two different gesturing orders systematically lead to different interpretations, or that 2) two different gesturing orders do not show any pattern in their interpretation. The hypothesis is that we will find 1), so that there is a preferred interpretation for the gesturing order belonging to the motion event (SOV), and a preferred interpretation for the gesturing order belonging to intensional events (presumably SVO). The expectation is that gesturing is a communicative task, and that the gestures are chosen in such a way that the translation process of the interpreter are taken into account.

2.4.3 Overall framework

Moving to the broader framework of evolution of language for a moment, remember that one of the questions was whether protolanguages have existed, and that gesturing is seen as a possible stage. Kirby then noted that intermediate stages in an evolutionary process are only plausible if there is a function for it. The same goes for a theory of language that supposes protolanguages: "An adaptationist programme for human language, such as Pinker & Bloom's ([17]) which stresses communication as the adaptive function of language, is strengthened if functional intermediates can be found" ([14]).

The view that emerges from the combination of the results of the production and interpretation experiments might be able to offer such a *functional* intermediate: if it turns out that different gesturing orders indeed lead to different interpretations, then communication is the adaptive function of language systems that gesturing offers a window for. If we assume that gesturing is an analogue for early language systems, and show that gesturing has a function, then we can draw inferences about the function of early language.

In the next part, the experiments are described.

Part II

Experiments

Chapter 3

Production

3.1 Main Production Experiment

3.1.1 Method

The participants were nineteen Dutch native-speaking students (seven male, twelve female) recruited from a class at Utrecht University. The participants were not familiar with any sign language. They received a few extra points for their grade as a compensation for participation.

The stimuli consisted of twenty four drawn vignettes on which a simple situation was depicted. Beforehand, we constructed the sentences that were to be used for the vignettes. These are shown in table 3.1. As can be seen, each sentence consists of a subject, verb, and object. For each verb, two versions exist, each having a different subject and object. For example, for 'throw' the versions A1a and B1a were created, with two different subjects ('girl' and 'woman') and two different objects ('guitar' and 'saxophone'). Half of the items are motion events, while the other half are intensional events. In the table, the *a*-items are motion events (for example, 1a, 'throw') and the *b*-items are intensional events (for example, 1b, 'hear'). The motion events were added as a check whether we could find the SOV order that was so apparent in other works.

Table 3.1: Items for the production experiment.

	A		B
1a	girl throws guitar	1a	woman throws saxophone
1b	girl hears guitar	1b	woman hears saxophone
2a	woman lifts saxophone	2a	a girl lifts guitar
2b	woman draws saxophone	2b	girl draws guitar
3a	man catches shoe	3a	gorilla catches ball
3b	man thinks of shoe	3b	gorilla thinks of ball
4a	boy shoots elephant	4a	a girl shoots horse
4b	boy dreams of elephant	4b	girl dreams of horse
5a	gorilla feeds cat	5a	man feeds bird
5b	gorilla looks at cat	5b	man looks at bird
6a	woman pushes horse	6a	boy pushes elephant
6b	woman looks for horse	6b	boy looks for elephant

The items were divided in two subsets of twelve items, as depicted by the two colors in the table. This way, we prevented the subsets from having sentences with the same subject and object (like A1a and A1b), and we made sure that motion events and intensional events occurred equally often. The participants were randomly assigned to either the first or the second subset, so that the first set was given to nine participants (two male, seven female) and the second one to ten participants (five male, five female).

As was said above, for each sentence a vignette was drawn (see figure 7.1 for some examples). Each participant watched the twelve items on a computer screen in random order, and was asked to make clear what was happening on the picture without speaking. The experimenter was sitting opposite the participant and was unable to see the vignettes. The participant watched the vignette as long as he wished, then clicked the screen to make the vignette disappear and start gesturing. When he was finished, he clicked again to see the next picture. The camera inside the computer recorded the participant's movements. Two coders independently coded the gestures. Participants were asked to sign a form that gave their permission to be recorded. An example of a vignette ('A gorilla feeds a cat') and a gesturing participant is shown in figure 3.1.

Figure 3.1: Screenshot of vignette and gesturing participant

The participants were also asked to give written descriptions of the vignettes they gestured, to make sure they understood the pictures correctly. This task was done *after* the gesturing, to ensure that the written descriptions would have no influence on the gesturing task.

3.1.2 Results

During the experiment, the researcher wrote down the gesturing orders that the participant used. Afterwards, a second coder watched the recorded video's and coded the gestures. The coders decided that if a participant made several attempts at gesturing a vignette, the last attempt would be coded. The agreement of codings was higher than 90%.

We also checked the written descriptions of the vignettes that the participants gave after the gesturing task. While all descriptions contained correct subjects and objects, the verbs were wrong in 10.5% of all cases. However, it was always the case that the participants chose a verb of the same kind that we had intended on the vignette. For example, if the original sentence was 'a woman pushes a horse', it was sometimes the case that the participant instead wrote 'a woman pets a horse', in which both 'push' and 'pet' are motion events. Similarly, 'looks for' was often substituted with 'thinks of', both intensional verbs. For this reason, we did not delete any codings from the results.

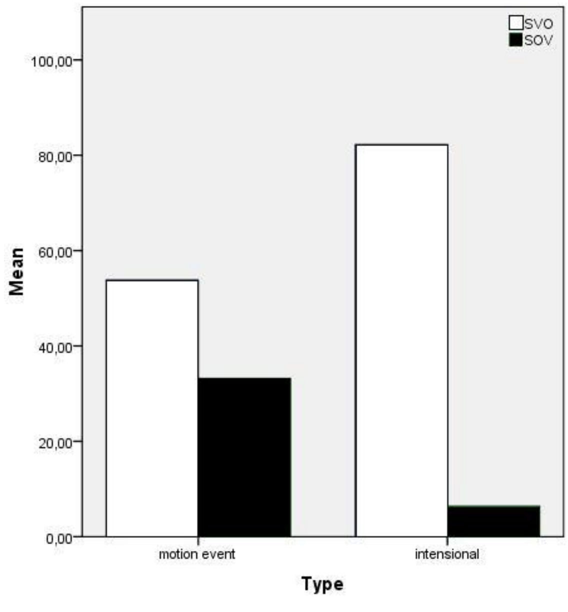
The main result of the experiment is shown in table 3.2.

Table 3.2: Complete overview of gesturing orders.

	One/Two-gesture strings 7.5%				Three-gesture strings 92.5%				
Type of sentence	OV	S	SV	VO	OSV	OVS	SOV	SVO	VOS
intensional	3	1	1	4	5		7	92	1
motion event	7			1	7	1	37	61	
Total	10	1	1	5	12	1	44	153	1

The table shows the complete overview of all the gesturing orders that the participants used. As was expected, SVO and SOV are the dominant ones: these constitute 86.4% of all coded results. The remaining orders are a mixture of two-gesture strings (such as VO and OV) and three-gesture strings (such as OSV and VSO). The two-gesture strings make up 7.5% of the results. In the remainder of this section, we will focus on the SVO-SOV contrast, and briefly get back to the two-gesture strings in the Discussion.

Figure 3.2: Average percentage of gesturing orders for the two types of items.



In figure 3.2, the average percentage of the gesturing orders SVO and SOV for the two types of items are displayed. On the left, we can see that motion event were gestured by the participants by using both orders, although SVO is more frequent. A paired t-test shows that this difference is not significant ($t(11) = 1.04$, $p = .32$). On the right, the intensional events show a more obvious difference: a substantial part of the participants chose to use SVO. A paired t-test shows that SVO is used significantly more often ($t(11) = 11.72$, $p < .01$).

If we compare the motion events to the intensional ones, there is a significant difference between both the white ($t(11) = -3.19$, $p < .01$) and the black ($t(11) = 3.05$, $p < .05$) bars. This means that SOV is used more often with motion events, and that SVO is used more often with intensional ones.

3.1.3 Discussion

As was expected from the findings in the pilot study by Schouwstra, the intensional items were gestured almost exclusively by using SVO order. This result is in line with the hypothesis that intensional verbs constitute a different kind of verb than the motion events. In terms of the conclusion that Goldin Meadow et al. have drawn, it seems to be the case that SOV is not a *universal* way of representing simple events.

A striking finding is that the motion events show no preference for a particular gesturing order. In line with the work by Goldin Meadow et al. and Langus & Nespor, we would have expected to find a clear majority for SOV gesturing. If this would indeed have been the case, our results for the intensional items would have been stronger, because it would show that our experimental setup was comparable to the work of the other researchers. As it is, there are several reasons that might explain the high numbers of SVO gesturing for the motion events.

The first and most obvious possible explanation is that the participants, who were all Dutch, used the ordering that they also used for the written descriptions, namely SVO. The descriptions consisted of a round 100% of SVO sentences. It could be that the intensional items triggered the grammatical system and thus the arising of SVO orderings, which also influenced the motion events. However, the descriptions were written *after* completing the gesturing task, so the expectation was that grammar did not have a role to play. There is no direct evidence that grammar is involved within gesturing, and in fact, we will argue later on that primarily the semantic properties are responsible for the various gesturing orders (and, as we will see, their interpretation).

In more general terms: if we assume that intensionality is in some way 'special' and triggers SVO, this could have influenced the gesturing of the motion events regardless of the explanation of the occurrence of SVO. The motion and intensional events were shown in random order during the experiment, which might have prevented the participants from 'getting into' the SOV order. Of course, we can not look at the raw data from Goldin Meadow and Langus & Nespor, but their results did not show a full 100% of SOV, which might indicate that it took the participants a few items to 'discover' that SOV was the most comfortable order to gesture the items. Another supporting piece of evidence for this line of thinking is that the pilot study by Schouwstra showed higher numbers of SOV orderings (75%), and in that study the motion events were grouped together, followed by a group of intensional items.

A last interesting possibility is that some of the motion events might have been intensional in nature. In table 3.3, the occurrences of SOV and SVO are listed for the motion events that were used in the experiment. Some of them show a clear preference for SOV, like A1a ('girl throws guitar'), while others show a mixture of SOV and SVO, like A6a ('woman pushes horse'). There are also items that were gestured solely using SVO, like A3a ('man catches shoe'). One could argue that some of the items that favoured SVO contain a verb that could be explained intensionally in combination with a particular object. For example, a man that is fishing does not know yet what kind of object he will catch, a fish or a shoe. The manner in which the participants gestured this particular item shows this intensional element: they gesture a man who is fishing, and who is quite surprised to find he has catched a shoe.

Table 3.3: Gesturing orders for the motion events.

Item	A1a	B1a	A2a	B2a	A3a	B3a	A4a	B4a	A5a	B5a	A6a	B6a
# SVO	1	1	2	1	8	9	8	8	7	9	4	3
# SOV	7	8	6	8	0	1	0	0	2	0	3	2

Furthermore, if we compare the results to those of Langus & Nespor, they obtain a quite high percentage of two-gesture strings (41.4% and 36.8% for the two groups of participants), while in this experiment it was quite low (7.5%). This may have something to do with the differing objects: for example, Langus & Nespor used 'boy catches *fish*', in which 'catches fish' could be generalized under one single gesture quite easily, namely a fishing rod (although this can not be checked because there is no access to the raw data). In contrast, catching a *shoe* is not something that is easily guessed from gesturing just a fishing rod. In other words, Langus & Nespor used verb-object combinations that have an object that is easy to incorporate ([5]). This may have led to the greater amount of two-gesture strings, and also makes the items more typical motion events. The high amount of SVO orderings for our items might be explained by the more intensional nature of the items that we originally intended to be motion events.

It seems then, that the absence of a clear SOV pattern for the motion events gives rise to different explanations. In an attempt to solve this puzzle, a control experiment was conducted that only used the motion event items from the current experiment. If the participants *did* show the gesturing order SOV when faced with only the motion events, it would give a strong indication that in this experiment, the intensional items had a strong influence on the motion events. A clear conclusion that *can* be drawn from this experiment, however, is that the intensional verbs show different behavior than the items used in other experiments.

The control experiment is described in the next section.

3.2 Control Experiment

3.2.1 Method

The participants were nine Dutch native-speaking students (all female) recruited from a class at Utrecht University. The participants were not familiar with any sign language. They received a few extra points for their grade as a reward for their participation.

The stimuli were a subset of those from the main production experiment, namely the motion events. Table 3.4 shows the descriptions of the vignettes of the motion events. Two of the items were used as practice material. For clarity, the numbering of the items is the same as in the main production experiment.

Table 3.4: Items for the production control experiment.

Number	Item
A3a	man catches shoe
A2a	woman lifts saxophone
A1a	girl throws guitar
B6a	boy pushes elephant
B1a	woman throws saxophone
B2a	girl lifts guitar
A4a	boy shoots elephant
A5a	gorilla feeds cat
A6a	woman pushes horse
B3a	gorilla catches ball
B4a (practice item)	girl shoots horse
B5a (practice item)	man feeds bird

Again, each participant watched the ten items on a computer screen and was asked to describe what was happening on the picture without speaking. This time, all the participants watched the items in the same order (the order of the list in table 3.4). The experimenter was sitting opposite the participant and was unable to see the vignettes. The participant watched the vignette as long as he wished, then clicked the screen to make the vignette disappear and start gesturing. When he was finished, he clicked again to see the next picture. The camera inside the computer recorded the participant's movements. Two coders independently coded the gestures. Participants were asked to sign a form that gave their permission to be recorded.

As before, the participants also gave written descriptions of the items after they had performed the gesturing task.

3.2.2 Results

The recorded clips were coded by two independent coders. The correspondence of codings was higher than 90%.

Again, we also checked the written descriptions of the vignettes that the participants gave after the gesturing task. This time, the verbs were wrong in 5.6% of all cases. All the errors occurred with the verb 'push', which was interpreted as 'pet' a number of times. Like before, because the mistakes were within the same category of the original item (motion event), we did not delete any data.

The main result of the experiment is shown in table 3.5 (*N.B.* B4a and B5a are practice items and are therefore missing from this table).

Table 3.5: Overview of gesturing orders for the control experiment.

	Two-gesture strings 28.9%				Three-gesture strings 71.1%					
Item	OS	**OV**	SV	VO	OSV	OVS	SOV	**SVO**	VOS	VSO
A1a		2		1		1		4	1	
A2a	1	1					1	4		
A3a		1				1	1	5	1	
A4a		3						4		2
A5a			1		1			7		
A6a		2		1		1		3		1
B1a		3						5	1	
B2a		3		1				4		
B3a					1			8		
B6a		4		2				3		
Total	1	**19**		5	2	3	2	**47**	3	3

The table shows the complete overview of all the gesturing orders that the participants used. In contrast to the main production experiment, the participants used a lot more two-gesture strings: 28.9%. If we look at the two- and three-gesture strings as two separate groups, it is surprising that both of them have a different majority. The two-gesture strings consist of 73.1% OV, while the three-gesture strings consist of 73.4% SVO.

In the main production experiment there was no in depth analysis of the two-gesture strings. Instead, the numbers of SVO and SOV were compared. In this case, SOV only occurred 2 times out of 90. It might therefore be wise to merge the two- and three-gesture strings, and compare the numbers of SOV+OV and SVO+VO. This is shown in table 3.6.

Table 3.6: Reduced overview of gesturing orders.

Item	A1a	A2a	A3a	A4a	A5a	A6a	B1a	B2a	B3a	B6a	Total
# (S)VO	5	4	5	4	7	4	5	5	8	5	52 (71.2%)
# (S)OV	2	2	2	3		2	3	3		4	21 (28.8%)

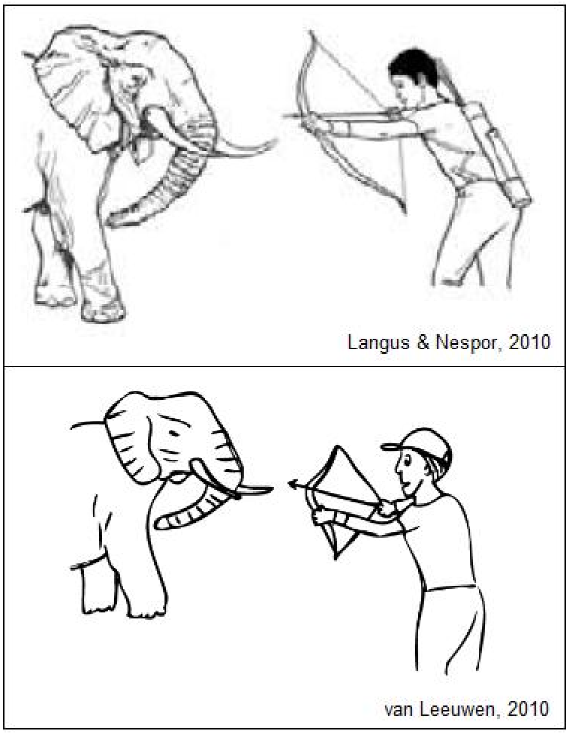
A paired t-test shows that the (S)VO ordering is used significantly more often than the (S)OV ordering ($t(10) = -3.59$, $p < .01$).

3.2.3 Discussion

The aim of this control experiment was to duplicate the results that Goldin Meadow and Langus & Nespor found in their production experiments. The expectation was that the motion events, without interference from intensional events, would predominantly be gestured using the SOV order. Obviously, the result that was obtained is the exact opposite: a clear majority of SVO.

This finding is surprising, mainly because the method was almost the same as that from Langus & Nespor. It was even the case that a lot of the vignettes are alike. For example, see figure 3.3, in which two versions of the vignette for 'boy shoots elephant' are shown: the top one by Langus & Nespor, and the lower one that was redrawn for this experiment.

Figure 3.3: Two versions of the vignette for 'boy shoots elephant'.



As was said, it would have strengthened the results of the main experiment if a SOV preference was found here. The clear dominance of the SVO order may have several causes. First of all, referring to the Discussion in the previous section, the explanations that were mentioned there could be applicable here as well. The possible interference of grammar, or the slightly intensional character of the items could have been the cause of the unexpected results.

Another factor that might have played a role is that the participants were drawn from the same group of students as the participants for the previous experiment. Since the lapse of time between the two experiments was about two weeks, chances are that the students talked about the first gesturing task to the students that were to participate in the control experiment. If they were aware of the fact that they would have to gesture sentences, they may have reflected on a strategy to do so, and have gotten a focus on spoken language that we were trying to avoid. Afterwards we talked to the participants and our suspicions were confirmed: some of them said they had 'written' sentences in mind while gesturing.

3.3 Overall Discussion

3.3.1 Comparison with other researchers

Goldin Meadow et al. were the pioneers of gesturing research, and the first to find the preference for SOV. Seated in a research area about homesign languages (the gesturing counterpart of pidgins), the conclusions were not quite strong or explicitly formulated in the framework of evolution of language. They state that the SOV order is "the one found in the earliest stages of newly evolving gestural languages and thus may reflect a natural disposition that humans exploit not only when asked to represent events nonverbally, but also when creating language anew".

The 'events' that they talk about are motion events, so it could be said that Goldin Meadow et al. did not get the whole picture, because they did not investigate intensional verbs. The conclusions they draw are not as 'advanced' as the ones by Langus & Nespor, who claim that improvised gesture communication is the product of a direct link between the conceptual and the sensory-motor systems in the brain, without the grammatical system interfering. As far as this claim goes, we agree that gesturing gives us a window into this 'grammar-less' state. A possible explanation for the 'failure' of the control experiment is that participants went out of this state, and let grammar back in, because they were consciously thinking about written or spoken language.

However, the assumption is that in the main experiment the participants indeed were only acting at the conceptual level. Furthermore, as was explained in Part I, it is also assumed that communication was acting as well. The results that were found mean that there is indeed a difference in motion and intensional events within the conceptual system, which resulted in the 'speaker' choosing different gesturing orders: SOV for motion events, and SVO for intensional ones. Langus and Nespor strictly convey SOV to the conceptual realm (with its counterpart Actor-Patient-Act), and SVO to the computational one (with counterpart Actor-Act-Patient), but that claim has now been weakened considerably. SOV is not the only structure that can be found when participants carry out a communicative, pre-linguistic gesturing task.

3.3.2 Multiple basic structures?

In Part I, it was already discussed that Goldin Meadow and Langus and Nespor place their conclusions at the cognitive level. They claim that the gesturing structure SOV originates from a structure in the brain, ArPA. This thesis so far contains no claims about cognitive structures, for reasons given earlier. It may be worthwhile to speculate along the lines of Langus and Nespor, however, and see how they might respond to the finding of SVO.

For example, they might say that they were wrong in saying that the conceptual system only knows the ArPA structure. Instead, there are two basic structures: ArPA and ArAP, which deal with separate kinds of events. ArAP can still be the unique structure for grammar, and thus partly overlap with the conceptual system.

However, if we must stay in the tradition of postulating a 'basic structure', it's more plausible to take another route. Instead of assuming the existence of a 'list' of basic structures, the other option is that there is a semantic interplay that determines the particular gesturing order. It could be that ArPA is indeed the basic one, but when the semantics of the situation change, the ordering is adjusted to deal with that. In this view, the semantic properties of a situation determine the way it is represented.

The reason that this view is more convincing is that there is also the result by Meir et al., who found that animacy of the object can influence the gesturing order. Instead of saying that there is yet another basic structure for (in)animacy, it seems more intuitive to say that it is a semantic force that also plays a part in determining the structure of a representation. Ockham's razor ([19]), the principle that entities must not be multiplied beyond necessity, is also on the side of the second option. It favors a simple explanation over a more complex one. Postulating multiple basic structures involves more entities than having a system of semantic interplay.

Thus, the relation between cognitive and gesturing structures remains unclear. The result from the production experiment at least tells us that SOV is not the only gesturing order, which gives rise to a number of alternative explanations regarding the cognitive story. We now turn to the next part of the experiments about interpretation.

3.3.3 Next question

In the production experiment, we found several ways of gesturing events, namely SOV and SVO. In the interpretation experiment, the task is to examine the other end of communication: the interpretation by the 'hearer'. The task is to examine whether SOV and SVO indeed have a communicative purpose: will those two orders lead to different interpretations?

Chapter 4

Interpretation

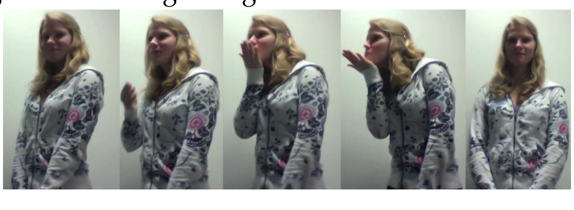
4.1 General Method

Unlike the production experiments, which were carried out using vignettes right away, the interpretation experiments slowly evolved by using different methods. The underlying idea, however, was the same for all experiments. Participants have to watch a movie of someone gesturing, and then decide what the gesturing meant. To investigate whether there are differences in interpretation between the SOV and SVO orders, several elements are needed.

- The stimuli: short clips of someone who is describing a simple situation using gestures. These movies have to be recorded in both the SOV and the SVO order, and then be shown to participants.
- Answer possibilities to choose from after the clip is finished. At the outset of the experiments, the idea arose to let participants give descriptions of the clips freely, e.g., without having any answer possibilities at all. They would have had to watch the clips, and then describe what they thought was being gestured. However, a quick test of this method showed that this was too hard for participants and led to quite a lot of different answers. Therefore, two or three answer options were required.
- A way to measure the responses. In the experiments, two measures are used: 1) Reaction time, and 2) *Kind* of response. The first one measures the time between the end of the gesturing clip and the moment the participants chooses an answer. The second one keeps track of what option was chosen. More details about that can be found in the descriptions of the separate experiments.

While all three elements are present in every experiment, they do vary in the way they are implemented. For the stimuli, there is a difference in whether the gestures that are used are ambiguous or not. When a gesture is ambiguous, it means it was recorded in such a way that two interpretations were possible. For example, in figure 4.1 a gesture is shown that was intended to mean both ’think of’ and ’kiss’. Detailed explanations are given in the sections below.

Figure 4.1: Ambiguous gesture for ’think of’ and ’kiss’.



Another varying element is the answer possibilities. Both written sentences and vignettes were used as answer options. In table 4.1 the three experiments are listed along with the particular method that was used. In the following sections, the method and results of the experiments are discussed.

Table 4.1: Overview of experiments.

Experiment	**Kind of movies**	**Answer possibilities**
1 (pilot)	Unambiguous	Written sentences
2	Ambiguous	Written sentences
3	Ambiguous	Pictures

4.2 Pilot

4.2.1 Method

The participants were fourty-two Chinese native-speakers recruited during a conference on linguistics. They did not receive any reward, but merely participated out of interest. The stimuli consisted of eleven gestured situations, all recorded in two gesturing orders, so that the total number of gesturing clips was 22. Beforehand, we constructed the sentences that were to be gestured in the clips. These consisted of the following: 4 motion events (including a practice item), 3 intensional events, 2 sentences containing a temporal adjective, and 2 sentences containing a negation. The sentences with a negation or a temporal word were used as fillers. The main items were clips containing three gestures, subject, object, and verb, recorded in two versions, namely SOV and SVO.

The answer possibilities were three written sentences for each item, which were translated to Chinese by a native speaker. There was always one option that was the same as the sentence we originally intended to be gestured in the clip. Next to that, the second option was a sentence of a different kind than the correct answer. For example, if the correct answer was a motion event, the second answer possibility would be an intensional event, and vice versa. The last option was a filler, and could either be motion event or intensional. Table 7.1 in the Appendix shows an overview of the original sentences and the generated answer possibilities.

The participants viewed the clips on a computer screen. When the clip was over, the answer options appeared, and the participant had to click on one of them. After that, the next movie appeared. Beforehand, a screen with instructions appeared. Also, the first item was a test item to let the participants get used to the set-up and the way the application worked. The items after the test item were shown in a random order. Also, for each item, the participant randomly received either the SOV or the SVO version.

In line with the findings from the production experiment and what we know from other research, the underlying thought here is that we expect the SOV order to facilitate interpretation of motion events, and SVO to facilitate interpretation of intensionality. We expected that when a motion event was gestured 'the wrong way', e.g. using SVO instead of SOV, participants would be more likely to choose the 'wrong', intensional answer. The same was expected for the intensional events. Furthermore, even if the participants were able to pick out the correct answer for both gesturing orders, we expected that they would do this more quickly for the correct ordering. For example, if everyone gave the correct answer for the first item, which is a motion event, we would expect the average reaction time to be lower for the SOV than for the SVO order.

The computer that was used to run the experiments kept track of the reaction times as well as the given answers.

4.2.2 Results

Out of 462 answers, 226 were correct (48.9%). Distributed over fourty-two participants, this means that the average number of correct answers was 5.4 out of 11 (including the test item). If the participants had chosen the answers in a random way, the expected number of correct responses would have been a third, because there were three answer options. As it is, they did slightly better than chance.

Upon closer inspection, the low ratings were caused by the last four items (the ones that contained a temporal aspect or a negation). The number of correct reponses for the test item and the first six items was 183 out of 294 (62.2%). In table 4.2, an overview is given of the responses for these items. The columns 'SOV' and 'SVO' indicate which order the gesturing clip was shown in, and 'RT-sov' and 'RT-svo' denote the average reaction times for the two orders. All grey columns show the correct answers, but the darker grey ones show the situation that was expected to be the most 'easy' for the participants. For example, the test item is a motion event, so it was expected that SOV would be the best gesturing order to convey that message. Among the answers for the SOV order, 63.2% chose the motion event answer, 36.8% picked the intensional answer, and no one chose the filler.

Table 4.2: Overview of results first interpretation experiment.

Item & Type	Type of answer	SOV (%)	SVO (%)	RT-sov (s)	RT-svo (s)
test - motion event	m.e.	63.2	87	6.5	14.4
	int	36.8	8.7	7.5	7.7
	filler		4.3		13
1 - motion event	m.e.	35	84.6	12.1	9.5
	int	35	7.7	15.4	1.8
	filler	30	7.7	16.2	13.2
2 - motion event	m.e.	16.7	26	12.2	16.2
	int		7.3		11.8
	filler	83.3	66.7	12.1	8.4
3 - motion event	m.e.	95	85.7	12.6	11.6
	int				
	filler	5	14.3	6.3	22.9
4 - intensional	m.e.	20	15.4	7.6	17.6
	int	60	50	11.2	11.1
	filler	20	34.6	15.8	8.6
5 - intensional	m.e.	4.8	4.8	12.5	14.5
	int	81	90.4	9.5	9.3
	filler	14.3	4.8	15.5	1
6 - intensional	m.e.	27.8	6.7	23	11.5
	int	50	73.3	14	14.6
	filler	22.2	20	11.4	11.2

If we take a look at the table, it is clear that the responses in general were divided between all three answer options. It is not possible to draw a firm conclusion from this table. Some items were answered fairly well by the participants (like 3 and 5), while others seem to have been answered randomly (like 4 and 6). The expectation that the 'wrong' gesturing order would lead to a wrong answer is not fulfilled: there is no correlation between *SOV - motion event answer* on the one, and *SVO - intensional answer* on the other hand.

Furthermore, the reaction times do not confirm the hypothesis either. For example, item 3 was answered correctly by almost all participants, but there is no big difference in reaction times for the SOV and SVO order.

To examine whether the cause of these confusing results was the difficulty of the items, a second analysis was carried out using only the participants who had gotten at least 5 items correct out of the test item + first six items. Again, the reaction times showed no significant difference between the two orderings ($t(6) = -1.1$, $p = .31$).

4.2.3 Discussion

This pilot study did not deliver any useful results. A number of issues in the set-up can be identified. First of all, since the given answers were distributed between all the options, it seems likely that three options is too much. After watching a gesturing clip, it takes a while to examine each answer possibility, in which time the effect of the SOV or SVO ordering may already have disappeared. Also, in hindsight the answers that were designed for this study may not have been very appropriate. After designing the main sentence that would be gestured, the other two answer options were created, which caused those options to be less fitting to the gesturing than the main sentence. It would have been better to first design answer options that are similar, and then come up with the gesturing clip (as we will see in the next two experiments).

Another drawback was that the answer possibilities were translated into Chinese, which meant for us that as an experimenter, there was no more control over them. The translator declared that she had to change a few words because they were simply 'untranslatable'.

The last factor that may have played a role is that the duration of the gesturing clips, which was about twenty seconds. In comparison to the work by Langus & Nespor, this is quite long, because their clips took only six seconds. Twenty seconds may have made the clips too difficult, or given the participants the time to start thinking about the gesturing too consciously. The filler items may have been too difficult altogether, which may have affected the participants while answering the 'real' items.

To summarize, then, the set-up of the interpretation experiment needed to be improved.

4.3 Experiment 2

4.3.1 Method

The participants were thirty Dutch native-speakers (ten male, twenty female) recruited in the Utrecht University Library. They did not receive any reward, but merely participated out of curiosity. After the disappointing result from the pilot study, the set-up of this experiment was revised. The gesturing clips were constructed in such a way that they contained an *ambiguous* gesture. As was explained in the section about the general set-up of the interpretation experiments, two versions were recorded for each ambiguous verb gesture, namely SVO and SOV, which we hoped would lead to be interpreted respectively as motion event and intensional event. For example, we started out with an ambiguous gesture X which could either mean 'kiss' or 'think of'. Then, the two versions that we recorded were 1) S-V-O, which should lead to 'think of', and 2) S-O-V, which should lead to 'kiss'. There was a total of 4 ambiguous clips, and two practice items that were not ambiguous. The filler items containing temporal words or a negation were deleted from the study.

For each ambiguous item, the two possible interpretations of the verb were the two answer possibilities, presented as Dutch written sentences. In the example above, both S-V-O and S-O-V were presented with the same two answer options. This way, the gesture-pairs could be compared for the number of motion event and intensional event responses. Table 4.3 shows the items that were used and the corresponding possible interpretations (N.B.: for the first item, "A woman paints an elephant", no appropriate translation is available in English. The two options that are meant here are "A woman paints the skin of an elephant" and "A woman creates a painting of an elephant").

Table 4.3: Overview of items for the second interpretation experiment.

#	Complete sentence	SOV order	SVO order
test	"A strong man throws a ball"	strong man - ball - throws	
test2	"A girl knits a scarf"		girl - knits - scarf
1	"A woman X an elephant"	woman - elephant - paints	woman - paints - elephant
2	"A girl X a book"	girl - book - sleeps on	girl - dreams of - book
3	"A king X a ball"	king - ball - drops	king - looks for - ball
4	"A woman X a strong man"	woman - strong man - kisses	woman - thinks of - strong man

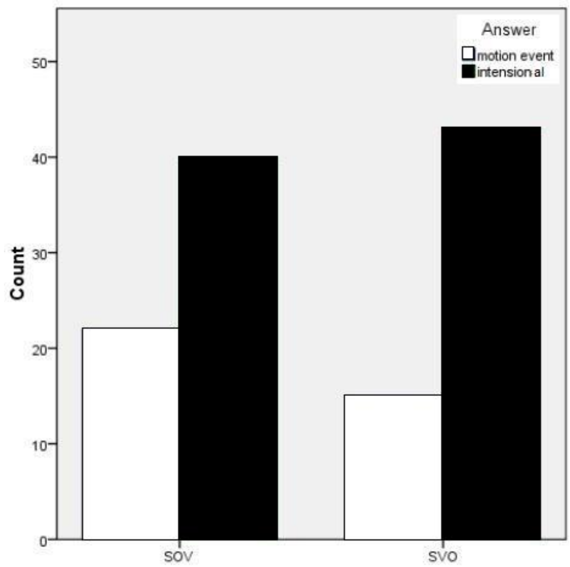
Like before, the participants viewed the clips on a computer screen. When the clip was over, the answer options appeared, and the participant had to click on one of them. After that, the next movie appeared. Beforehand, a screen with instructions appeared. The first two items were test items, and were recorded in one order only (see table 4.3). The items after the test items were shown in a random order. Also, for each item, the participant randomly received either the SOV or the SVO version.

The same application was used as before, and it kept track of the reaction times as well as the given answers.

4.3.2 Results

For the gesturing clips in the SOV order, 64.5% of the participants chose the intensional option (40 out of 62 responses), while the other 35.5% picked the motion event answer (22 out of 62). For the SVO order, these numbers were almost the same: 74.1% intensional (43 out of 58) and 25.9% motion event (15 out of 58). Figure 4.2 sums up these results.

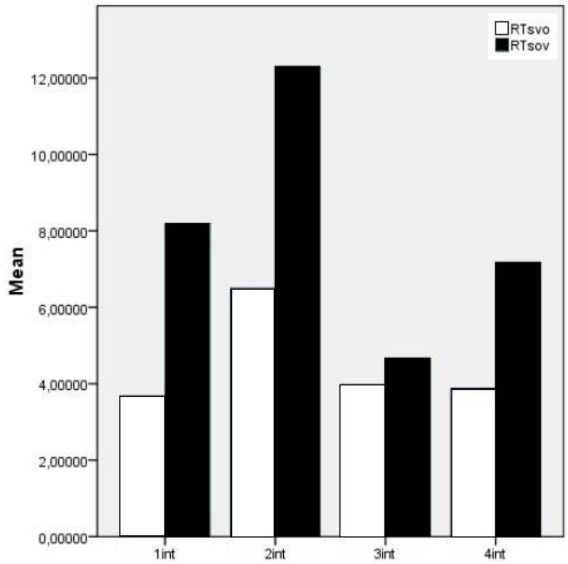
Figure 4.2: Overview of results second interpretation experiment.



Within the SOV order, the difference between the two kinds of responses is not significant ($t(3) = -2.14$, $p = .12$). For the SVO order, the difference *is* significant ($t(3) = -3.74$, $p < .05$). There are no significant differences between SOV and SVO order (for the white bars $t(3) = 1.33$, $p = .28$, and for the black bars $t(3) = -.68$, $p = .55$).

For all individual items except one, the number of intensional answers was bigger than the number of motion event answers. Clearly, there was a general preference to pick the intensional answer. For the SVO order, this was expected, but not for SOV. Therefore, the reaction times were analyzed for the situations that an intensional answer was picked. In figure 4.3, the horizontal axis shows the four intensional answers (one for each of the four ambiguous items), and the vertical axis shows the separate mean reaction times for when the item was presented in SOV or SVO order. As can be seen, the black bars are higher in all cases. This means that it takes longer to pick the intensional answer when the participant just saw a clip in SOV order, than it does when the participant saw SVO order. In fact, this difference is significant ($t(3) = -3.29$, $p < .05$). For the answers of type motion event, there is almost no difference in reaction times for the two orders ($t(3) = -.56$, $p = .62$).

Figure 4.3: Reaction times for intensional answers.



4.3.3 Discussion

The hypothesis was that the SOV order would trigger motion event answers, while SVO would lead to intensional answers. The prediction for SVO has been correct, because there is a significant preference there for the intensional answers. However, this pattern also extends to the SOV order, which was highly unexpected. In particular, if we connect the results from the production experiment to this study, the high number of intensional choices can not be accounted for.

Recall that in the production experiment, we saw that motion events were gestured using both SVO and SOV (there was no significant difference), and that intensional items were gestured almost exclusively using SVO. Thus, SOV only appeared for the motion events, and we would predict that in the interpretation experiment, SOV would solely lead to an interpretation of a motion event. This is not what happened. The black bar on the side of SOV in figure 4.2 is not explained: it would suggest that intensional items should be gestured with SOV during the production experiment as well, which was clearly not the case.

On the other hand, the white bars *do* fit the data from the production experiment. Since motion events were gestured using both SVO and SOV, we could have expected that both gesturing orders would be able to lead the participants to interpret them as a motion event, which they did in 35.5% of the SOV clips, and 25.9% of the SVO clips.

While we were interested in the kind of responses the participants would give, the only significant and exciting result is that of the reaction times. The reaction times that were analyzed in the results section suggest that when participants were presented with SOV order, they were heavily in doubt about what option to pick, whereas the SVO order quickly led them to the intensional answer. This is an indication that the plan to lead the participants toward a particular answer almost worked: they were having trouble choosing an intensional answer when confronted with SOV, but in the end still picked this option. This shows that the set-up of the experiment could still use one more improvement so that the participants would no longer hesitate to pick a motion event when they watched SOV.

Of course, the big difference between the production and the interpretation experiment is that in the former, vignettes were used as stimuli for the participants, and in the latter, we used written sentences as answer options. To have a perfect 'mirror' experiment, we should offer vignettes to the participants in the interpretation study as well. That way, the production experiment is characterized by *vignettes* → *gesturing*, and the interpretation experiment is characterized by *gesturing* → *vignettes*.

In hindsight, it makes sense to *not* use written sentences for answers, since these studies are intended to be a window on language evolution. Keeping written and spoken language out of the picture is a good thing if we want to investigate the more primitive system of gesturing. Langus & Nespor already used this insight, and in the third and last interpretation experiment of this thesis, it will be used as well.

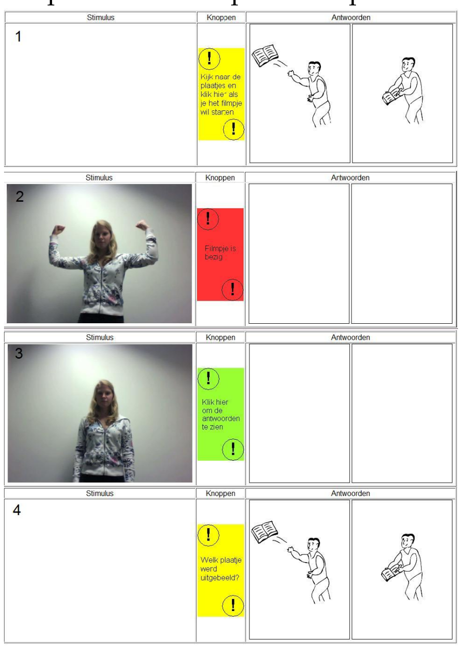
4.4 Experiment 3

4.4.1 Method

The participants were twenty-three Dutch native-speaking students (eight male, fifteen female) recruited in the Utrecht University Library. They did not receive any reward, but merely participated out of curiosity.

The set-up of the experiment is shown in figure 4.4, in which the first test item is playing ("A strong man throws a book").

Figure 4.4: The experimental setup for interpretation experiment 3.



The gesturing clips were the same as in the second experiment: four ambiguous clips, and two practice items (see table 4.3). This time, the answer options were presented as vignettes instead of written sentences. For each pair of SOV- and SVO gesturing clips, two vignettes were drawn, corresponding to the written sentences that were used before. See figure 7.2 in the appendix for the vignettes.

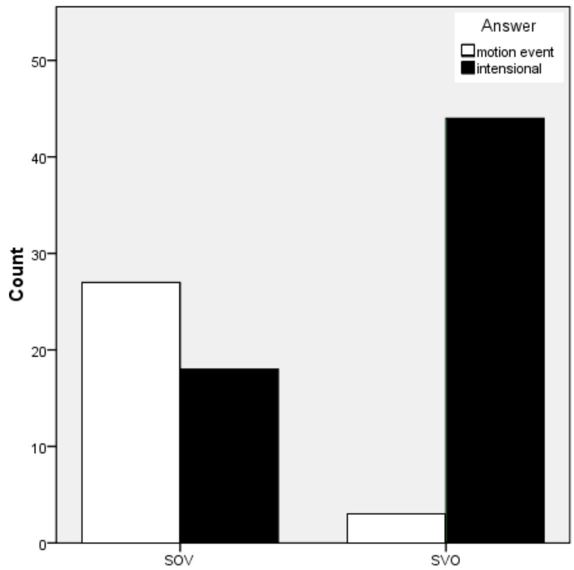
The experiment was carried out as follows: before the gesturing clip was shown, the participant looked at the two vignettes that he would have to choose from later on (step 1 in figure 4.4). The instructions explained that once the participant had examined the pictures closely, he could click the button and a short clip would play (step 2). Afterwards, after clicking again (step 3), the vignettes appeared again, and the participant's task was to choose which one was being gestured in the clip he had just seen (step 4).

Like before, the application kept track of the given answers and the reaction time, and furthermore, the *waiting time* before the gesturing clip starts was recorded as well. That way, it was measured how much time the participants took to study the vignettes before moving on to the clip.

4.4.2 Results

The main results are presented in figure 4.5. On the SOV side, the picture is quite different from the result of the second experiment. Of all the occurrences of the SOV order, 60% was answered with a motion event vignette, and 40% chose an intensional vignette. The SVO occurrences were more clearly divided: 93.6% intensional answers, and 6.4% motion event.

Figure 4.5: Overview of results third interpretation experiment.



If we compare the two orderings, both the white and the black bars differ significantly (white bars $t(3) = 6.06$, $p < .01$, for the black bars $t(3) = -6.06$, $p < .01$). Clearly, within the SVO order, there were significantly more intensional responses ($t(3) = -5.31$, $p < .05$).

The distribution within the SOV order is still not quite as clear as was expected. There is no significant difference between the two kind of responses ($t(3) = 1.52$, $p = .29$). Since the number of items is not that high, the distribution might be steered by an individual item. In table 4.4, the kind of responses are listed for each separate item. There is a clear pattern; for all the *a*-versions, more motion event responses were given, and for the *b*-versions, the intensional ones are in the majority. There is one exception: item 1*a*. This shows that on average, there was a clear distinction within the motion events between SOV and SVO, but the result was somewhat blurred by the first item. This may have had to do with the improbability of a person painting on an elephant, which was the motion event of item 1a (see also the pictures of the vignettes in the appendix, figure 7.2).

Table 4.4: Overview of results per item.

	Item								
	1a	1b	2a	2b	3a	3b	4a	4b	Total
Motion event reponses	5		8		6		8	3	30
Intensional reponses	8	10	3	12	3	14	4	8	62
Total	13	10	11	12	9	14	12	11	92

The average waiting time was 7.6 seconds, and the average reaction time was 4.3 seconds. An analysis of the reaction time per gesturing order shows there were no differences like the one we found in the second study. If participants picked an intensional answers, they did this equally quick for the SVO order as for the SOV order. The same goes for motion event answers.

4.4.3 Discussion

The third and last interpretation experiment was more successful than the first two, and delivered the results we had predicted early on in this thesis. The results show that we have found a pattern of interpretation: SVO leads to an intensional interpretation, while SOV will trigger a motion event response. In the general picture, this fits in with the view of communication that we have sketched earlier: the speaker will consciously pick an utterance order that will ensure the message gets across to the hearer. Our findings are quite impressive, given the opinions of other researchers on this subject.

This has probably to do with the improved set-up: by letting the participants study the vignettes before they saw the actual gesturing, they had a better sense of what they were supposed to watch out for. The waiting time confirms this: on average the participants took 7.6 seconds to look at the vignettes before they clicked to go to the gesturing clip, which is quite long. The average reaction time is only 4.3 seconds, which is quite short when compared with for example the reaction times for the first interpretation study (see table 4.2). This shows that once the participants knew what they had to look for, they were able to give a quicker response.

The results are strengthened by the fact that we used the same verb-gesture for both SOV and SVO, because it takes away the chance that the gestures in themselves contained a bias towards one of the two interpretations. The interpretation is only steered by the ordering of the gestures, which is exactly what we were trying to investigate. The interpretation study by Langus and Nespor used two vignettes as answer options that differed from each other in a random way - either the subject, the object, or the verb in the picture was different. This made the experiment very susceptible to the danger that the gestures in themselves led to a particular interpretation, and not the ordering of the gestures. Indeed, their participants reached a high percentage of correct responses, which underlines our thought.

A drawback of this study is that the number of ambiguous items is quite low (a bigger number would have made the results stronger). The explanation for this is that it is quite hard to construct such items: one has to come up with a gesture that resembles both a verb that is intensional and a verb that is a motion event. In the future, we might think about using 'artificial' ambiguous gestures, i.e., gestures that are random movements, and train participants to let them remember two meanings. Afterwards, if we show those gestures in both SOV and SVO order, the set-up of this experiment can be re-used. For now the results are very promising.

4.5 Overall Discussion

4.5.1 The three interpretation experiments

There is a tendency of improvement in the way the three interpretation experiments were carried out, which shows itself in the improved results. The first experiment suffered from too much complexity: the participants had to read three answer options, which were all quite similar, and then think back to the gestures they just saw. This was too difficult, and as a consequence the results were very diffused.

The second experiment went better, using only two answer options, and switching to the set-up in which ambiguous items were used. However, there still was the missing ingredient of using vignettes instead of written sentences as answer options. Reading the answers still made the task too confusing for the participants, because they first had to read and compare the two options and then decide on which one they would pick.

This was resolved in the final experiment, because the answers were now visualized with vignettes. The complexity of having to read the written answer options and relate them to the gesturing was removed, because the answers were much easier to comprehend than before. Also, using vignettes fits better with the distinction we made between cognitive structure and utterance structure. Written sentences have a linear structure, while thoughts presumably do not. To test the participant's interpretation, which is about thoughts, in hindsight it was obvious to use the vignettes. Another improved aspect was to show the vignettes before the gesturing, so that the participants knew what they had to focus on.

In the end, after a number of improvements, we got the impressing results we predicted in the earliest stages of the research.

4.5.2 Communication

Our last two interpretation experiments were truly unique because they used the same two answer options for the SOV and SVO orderings containing an ambiguous verb-gesture. In the study by Langus and Nespor, the answer options consisted of two vignettes that randomly differed in subject, object, or verb. One of those was the correct answer, and the other was wrong. They then shuffled the gesturing clips to create all six possible orders using S, O, and V, and showed these to the participants. It looks like the same principle, but it is not, because they presupposed that *whatever gesturing order was shown, the participants would choose the correct answer*, only the speed of the response would differ. What was done in this study, however, was that there was *no correct answer*: we were only interested in the *kind* of answer the participant would pick, *driven by a particular gesturing order*. This clearly shows Langus and Nespor's view: that there is only one way of representing events, and that the interpreter will get the message, no matter what gesturing order is used. In this thesis, it is argued that the gesturing order does matter, because of the message that is being conveyed. And indeed, the results show that there is a consistent pattern: SOV leads to motion event interpretation, and SVO to intensional interpretation. It is now clear that communication indeed is important here. The one that is gesturing chooses the gesturing order in such a way that the translation process of the interpreter will (in most cases) not fail.

4.5.3 Explanation of the 'coding pattern'

It was argued that the introduction of more than one gesturing order forces us to look at gesturing in a communicative context. The particular ordering that is chosen by the 'speaker' is chosen in such a way that the message will become clear to the 'hearer'. But why does SOV code motion events, and SVO intensional ones, and not the other way around?

It seems that the ordering varies according to the semantic properties of the event that is being described, or, as we termed it earlier, the thematic roles of the various constituents. There is a difference between motion events and intensional events in the way the object and the verb are connected to each other. For motion events, both the subject and the object are concrete entities that are 'present' in the here and now, and the verb is the link between them. Intensional events, on the other hand, involve a dependency of the object on the verb: the verb causes the state of the object. For example, in a situation where a subject is 'painting', the state of the object is not yet fully determined. Such conceptual considerations could be the explanation of the structures SOV and SVO. Langus and Nespor also consider this possibility (although obviously they only discuss SOV): "Gentner and Boroditsky (2009) have argued that relational terms - such as verbs - require the presence of the entities they link - such as nouns; suggesting that the SOV order may originate from the requirements imposed by the semantic relations in the conceptual system of grammar." Those semantic relations can be read as *thematic* relations, like we called them earlier.

4.5.4 Speculation about cognitive structures

As was the case in the discussion about the Production experiments, the view on the cognitive structures underlying gesturing orders remains unexplained. But again, it is worthwhile to speculate. Goldin Meadow and Langus and Nespor state that SOV is a linear cognitive structure ArPA, not just a gesturing order. If we follow their lead and state that both SOV (ArPA) and SVO (ArAP), and possibly other structures, are cognitive structures, and also accept our finding that gesturing orders definitely have a communicative function, that gives us another question. Did communication drive the development of cognitive structures for representing events, or did the cognitive structures develop on their own and thereafter influence communication?

According to the first option, the cognitive structures were formed under influence of communicative pragmatics. In the theory by Pinker and Bloom, it is assumed that the language faculty adapted to communication by natural selection. So, if you happened to have a particular way of structuring events, this gave you an advantage because you were better able to communicate. This way, cognitive structures evolved driven by communication.

The second alternative is that the conceptual structure already existed, and that communication by gesturing was simply following the orders that already existed. Those orders might be guided by semantic principles that were discussed in the previous paragraphs. In line with this second option, one could say that until language evolves into a more complex system, it is 'handy' to use the conceptual structure for your utterances. After all, as long as no agreed upon system exists, why not just use the structure that comes natural to everyone? Goldin Meadow agrees with this view: "we suggest that, initially, a developing language co-opts the ArPA order used in nonverbal representations and uses it as a default pattern, thus displaying SOV order, which may have the virtue of semantic clarity".

What view do our results support? After carrying out the interpretation experiments, it should be clear that the role of communication is not to be underestimated. It was shown that the one who is gesturing chooses an order that will make sure that the interpreter will translate it back to the right thought. The exact structure of the thought was left undefined in our discussions so far. A possibility is that the one way that the speaker could ensure that the message would be translated correctly, was to use the cognitive structure that underlies his message. So, to convey a message that contains an intensional event, he would pick SVO, because the interpreter was bound to translate that structure to the cognitive structure, ArAP. You could say that it is communicatively useful to choose the structure that is shared by everyone for representing events when you want to convey a message.

This view combines the thought that gesturing orders have a corresponding cognitive structure with the view that the choice of a gesturing order is driven by communication. However, incorporating earlier discussions, the thought that each gesturing order has a 'basic cognitive structure' counterpart is not very appealing. There has to be a point in which the rudimentary representation of a event, with all its semantic properties, is translated into a linear structure. Postulating a basic linear ArPA and ArAP structure will not solve that translation process, on the contrary, it takes the problem one level higher, because then where did ArPA and ArAP come from?

4.5.5 From semantic constraints towards grammar

The view that we like to defend here is that the representations that the participants had to convey during the gesturing task were, if anything, most similar to the vignettes that they were shown. Thus, the input for the gestures was a complex representation containing all the semantic information about the event. The semantic forces then drove the participants to choose a particular gesturing order that would ensure that an interpreter would be able to lead it back to the original vignette.

We have shown that there indeed was a communicative system within the gesturing orders. SOV consequently led to a motion event interpretation, while SVO led to an intensional interpretation. This system thus already contains some 'rules': "if the mental representation is intensional, preferably use SVO". Placing this in the overall debate about the evolution of language, it could be that early - spoken - language was guided by the same principles as was the gesturing in these experiments. Semantic properties of a thought determined the order of the words/sounds, which were then improved by pragmatic and communicative factors.

When the point came that the expressions became too complex, that is when grammar started to evolve. We thus propose a *gradual* view of the evolution of language, in which the computational system slowly evolved from the conceptual one. Contrary to Chomsky, there was no sudden revolution from which point grammar existed.

Part III

Final Conclusion

Chapter 5

Conclusion

5.1 Research Area

5.1.1 Paradigm

This thesis is seated within the framework of the evolution of language. The questions that have been discussed are concerned with a particular area within this field, namely the theories about the development of spoken language in its earliest stages. The existing debate is about whether language spontaneously emerged at a sudden point, or whether there was a gradual development.

The first insight that we discussed was the fact that there is a lack of direct evidence to work with. How can we study a process that took place thousands of years ago, and which left behind almost no traces? The method that we employ to overcome that obstacle is the Windows Approach: using 'living fossils' to draw inferences about earlier stages of language evolution.

The window that was used in the experiment is gesturing. Foremost considered to be a possible predecessor of spoken language, that image changed when Goldin Meadow and her colleagues let participants carry out a gesturing task. She found that when people gesture a simple event, they tend to do this using the order of subject-object-verb (SOV). The result was interpreted as signifying a cognitive structure ArPA for representing events that is independent from grammar, and thus a conceptual construct.

Langus and Nespor continued this work and added another component: not only letting people gesture, but also *showing* them clips of other people gesturing. Their research seemed to confirm the conceptual ArPA structure, because in the interpretation experiment, SOV was found as the preferred order of watching gesturing clips. Langus and Nespor then take their claims one step further, and make a sharp contrast between the conceptual system, which favours ArPA, and the computational system of grammar, which is connected to ArAP.

At this point of the discussion, our research began. The stimuli that the researchers used so far only included so called motion events: an event which involves a motion, be it in or across space. Influenced by the work of Schouwstra and Meir et al., we became aware that other types of stimuli generate different gesturing behavior. In the first set of experiments, we examined stimuli that contained an intensional item instead of a motion event, and repeated the gesturing tasks described earlier.

After that, we focused on the role of communication within the gesturing task. While Goldin Meadow simply denies that communication has any role, and Langus and Nespor also remain of the opinion that gesturing orders only have a cognitive basis, we argue that the influence of communication can not be denied. To show this, we set up a number of interpretation experiments. Following Langus and Nespor, participants watched a gesturing clip in various gesturing orders, and then chose between two pictures which one represented the situation that had just been gestured. The new element was that we used ambiguous gestures, that, when shown in a different order, would presumably lead to different *kinds* of interpretations. Langus and Nespor assumed that participants would always choose the correct answer, only the speed of the response would differ. What was done in this study, however, was that there was *no correct answer*: we were only interested in the *kind* of answer the participant would pick, *driven by a particular gesturing order*. This clearly shows Langus and Nespor's view: that there is only one way of representing events, and that the interpreter will get the message, no matter what gesturing order is used. But we were convinced that the gesturing order would matter, because it was picked by the speaker to convey a particular message.

5.1.2 Results

The production experiments, in which participants were asked to gesture a combination of motion and intensional events, showed mixed results. The clear *motion event - SOV* relation could not be duplicated. We hypothesized that maybe the intensional items influenced the participants, because for these items there was a clear majority of SVO gesturing orders. Thus, we found that contrary to other researchers' results, there were multiple gesturing orders at work here. By semantically adapting the stimuli, the participants responded by adapting their gesturing behavior.

On top of that, the interpretation experiments also provided new insights. If we take three gestures for a subject, an object, and a verb, the last of which could be interpreted in two ways on its own, and show these gestures in both the order of SOV and SVO, we found that participants chose different kinds of answers according to the order they saw the gestures in. SOV led to more answers of motion event, while SVO made participants pick the intensional answer. This confirmed our hypothesis that gesturing orders have a communicative function: the speaker can use them to deliberately influence the interpretation that the hearer forms.

5.2 Looking through the gesturing window

We used the gesturing window to take a look at the evolution of language. More specifically, we used gesturing to simulate a situation in which people were forced to communicate in a improvised way, unable to use their first language. It was intended to be an analogue for a situation in which language is emerging where no language existed yet. What do the results of the experiments say about this? In this section a general picture will be drawn about early language, parts of which are speculation, but overall guided by the obtained results.

In the early situation when language is slowly developing, the speaker wants to convey a message to a hearer. The message is quite simple, consisting of a subject, an object, and a verb. However, although the message is simple, some confusion may arise for the hearer, so the speaker needs to have a strategy to make sure the hearer interprets the speaker's utterance in the correct way. How can the speaker accomplish this task? Seeing as there is no computational system existing yet, the speaker only has semantic and pragmatic constraints to go on.

For example, the following three methods have been described during these thesis:

- *The verb*: If the message contains a relatively simple motion event, and the second argument is an inanimate object, then it is safe to use an ordering of subject-object-verb in your utterance. This follows the convenient way of first naming the arguments, and then the relation between them. If the verb is of another kind, and expresses an intensional event, the thematic relation between the verb and the arguments is different: the speaker had then better use subject-verb-object to denote that the object is dependent on the verb.

- *The (in)animacy of the object*: If, however, the event involves two animate arguments, the hearer might get confused about who is the actor, and who is the patient. The best solution now is to put the patient after the verb, so that there is a clear distinction between subject and object. Thus, the speaker will use subject-verb-object.

- *Relation between verb and object (incorporation)*: Sometimes the verb and the object are paired in such a way that the verb almost automatically 'implies' the object. For example, if the speaker wants to inform the hearer of the fact that he wants to catch a fish, then the verb 'fishing' and the object 'fish' are tightly related already. The verb has, in this particular case, 'swallowed' the object, and it would be enough to only utter subject-verb.

More may exist (we will come back to that in the chapter about future research), but the general thought is that thematic relations within the message will be used to convey a message. As long as no grammar is there yet (in the state the participants were during the production experiment), the forces at the conceptual level are all the speaker can work with to ensure a correct interpretation in the hearer's mind.

Thus, as long as the message was relatively simple, this system of communication could work quite well (as we have seen in the interpretation experiments). The semantic and pragmatic constraints were enough to keep communication going. It is clear that here, we can already speak of a *system*, which slowly deviates from being mere improvisation. As more complexity arose in the communicated messages, slowly also grammar developed. In a gradual way, the system based on semantics was no longer sufficient to handle communication, and gave rise to grammar. In this view, the division between the conceptual system and the computational one is not very strict, because the latter evolved out of the former.

Thus, the results of the two sets of experiments - production and interpretation - were combined to form this coarse picture, which of course needs a lot of additional research to complete it (see the chapter on future research). The production experiments helped to show that there is a semantic interplay at the conceptual level, and not just one structure. The interpretation experiments helped to show that gesturing indeed is a communicative task: by varying gesturing orders, the one who is gesturing can influence the hearer's interpretation. We have argued against the picture sketched by Langus and Nespor, in which SOV (ArPA) and SVO (ArAP) were confined to two separate modules of language, and in which grammar knew a sudden development. Instead, we propose that the conceptual system is not limited to one structure only, and moved the focus of research to the structure of the utterance. Also, we defy the view of a sudden language evolution, and instead argue for a gradual development.

5.3 Relation to CAI

During the discussion of the experiments and the literature that is involved, we have come across a lot of subjects of interest to the field of Artificial Intelligence. At the start of the thesis this was already pointed out, and here at the conclusion we briefly return to this point.

Within intelligence, language plays a large role. It is a challenge to transfer this system to computers, and to design computers that show human behavior - including communication. The experiments in this thesis have evolved around questions surrounding the nature of mental representations, the beginnings of spoken languages, and strategies of communication. It has thus not only been a linguistic, but also a cognitive debate. Of course, also the modules view and the contrasts or overlap that may exist between modules is an issue of importance to AI. Intelligence is primarily seated in the brain, so the question of what brain area accounts for a particular function, and the question whether there is such a strict division at all, are central questions.

The results of the experiments here thus not only give rise to new research questions within the framework of evolution of language, but also in the study of Cognitive Artificial Intelligence. In the last chapter, suggestions are given for further research.

Chapter 6

Future Research

In this final chapter, future research is proposed, both for the purpose of backing up the claims of this thesis, and for new areas that can be explored. It is by no means an exhaustive list, but it gives a guideline for additional studies.

6.1 Duplicate SOV result

First of all, the results that were obtained in the control production experiment contradict the claims of Goldin Meadow and Langus and Nespor. While the participants were only shown motion events, they still preferred to use SVO. In the discussion afterwards, it was already suggested that the participants may have been too well informed about their task, which could explain the unexpected results.

In any case, it is wise to design another control experiment that uses only motion events. This will strengthen the results from the main production experiment, because it shows that the general set-up was not flawed. Furthermore, it could also refine the set-up of the experiments. Maybe there will be a new insight into what particular aspect caused the participants to use the unexpected SVO instead of SOV.

6.2 Examine both SVO and SOV languages

The view that was presented here is that in the early stages of language development, semantic properties of a thought determined the order of the words/sounds, which were then improved by communicative and pragmatic factors (such as the context of the utterance, knowledge about the hearer, and the intent of the speaker). This system gradually evolved into a grammatical system, because the expressions become too complex to be handled by an improvised system alone. This view is in line with theories proposed by Jackendoff ([12]), who also claims that semantic and pragmatic 'forces' determined early language without the need of grammar.

However, there is one essential piece of evidence missing from this view, which was the power of the article by Goldin Meadow that started it all. Namely, the fact that Goldin Meadow et al. obtained their result from two groups of participants that differed in the predominant order of their mother language. This was the insurance that grammar was indeed being bypassed. Since in this thesis, only Dutch participants took part in the experiment, there is still one alternative explanation possible.

If indeed semantic constraints can only handle language up to a certain point of complexity, then it could be the case that intensional items were already too complex. In that case, gesturing intensional items would mean there was no longer a grammar-less state, and that the task was handed over to the computational system. The participants being Dutch, which follows SVO for simple transitive sentences, the gesturing would then lead to SVO, which is what we observed.

This could be explained by the motion events in a sense being more 'primitive' because they require only two modules to work together: the conceptual and the articulatory system. Intensional events might be more complex because they require the ability to think about objects that do not exist (yet). This complexity might have required the additional 'power' of a computational system of grammar.

To exclude that possibility, we would need to repeat the experiments with participants from a SOV language. If they gesture intensional items using SOV, it would point to an interference of grammar after all, because then the SVO speakers (Dutch) have used the order of their native language, and the SOV speakers as well. If instead they gesture intensional items using SVO, we obtain the piece of evidence that shows there was no grammar involved.

6.3 Hunt for other semantic influences

In this thesis, we have come across the three semantic factors that may influence gesturing behavior, which were summarized in the section above. As has been emphasized throughout the text, a simple situation consisting of a subject, an object, and a verb, can be semantically modified in a number of ways. While three of those ways have been described, it is not unconceivable that even more modifications may be identified that will influence the way an event is communicated. In the future, gesturing may be used to experiment with other types of events to see whether communicative behavior changes.

If the experiments are carried out with participants coming from the two most prominent languages in the world, SOV and SVO, the influence of grammar can be constantly checked. The question is at what point grammar does step in, i.e. how much can be seen through the gesturing window? Schouwstra in her work has already experimented with adding temporal adverbs to the simple events (some of which were used as fillers in the pilot interpretation experiment, as the reader may recall). Also, negation may be an interesting addition to study. Of course, temporality and negation are not quite 'simple' any more, so they may lead to the start of a whole new area of research.

6.4 Role of gesturing in evolution of language?

The window that gesturing offers was used here as an analogue window: the goal was to simulate a situation in which improvisation plays a great role, and in which there is no possibility to fall back on your first language. We have then tried to relate the findings offered by this window to the evolution of language, by discussing possible scenario's in which people first started to utter simple expressions.

However, as was described, gesturing is also a candidate for being a 'real' stage in the evolution of language, namely as a predecessor of spoken utterances. It is questionable whether we can ever find a window into that situation, after all, participants are equipped with the biological mark up that allows them to use language, which may not have been the case back then.

All in all, there is enough material to continue this interesting line of research.

Part IV

Appendices

Chapter 7

Additional data from experiments

7.1 Production Experiment

Figure 7.1: Examples of vignettes that were used for the production experiment. Intensional items are shown on the left, motion events on the right.

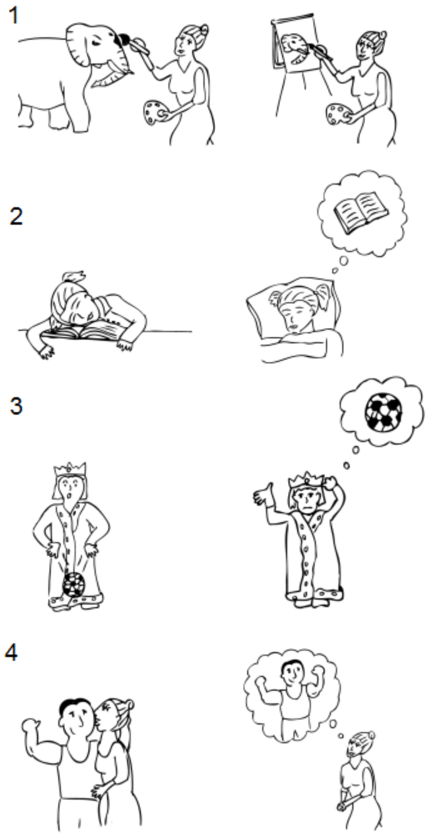
7.2 Interpretation Pilot Study

Table 7.1: Overview of items and answer possibilities for the pilot interpretation study (experiment 1).

	Item	MC options	
test	"A bodybuilder throws a ball"	#1	A bodybuilder throws a table
		#2	A bodybuilder breaks a record
		#3	A bodybuilder throws a ball
1	"A woman kisses a superhero"	#1	A woman kisses an officer
		#2	A woman kisses a hero
		#3	A woman thinks of a hero
2	"A waiter drops a spoon"	#1	A waiter looks for a spoon
		#2	A waiter eats with a spoon
		#3	A waiter drops a spoon
3	"An athlete shoots a cat"	#1	An athlete sculpts a cat
		#2	An athlete shoots a cat
		#3	An athlete is shot
4	"A woman knits a scarf"	#1	A woman knits a scarf
		#2	A woman lifts a scarf
		#3	A woman is cold
5	"A window cleaner hears a guitar"	#1	A window cleaner is tired of playing guitar
		#2	A window cleaner hears a dog
		#3	A window cleaner hears a guitar
6	"A carpenter looks for a hammer"	#1	A carpenter drops a hammer
		#2	A carpenter hurts himself with a hammer
		#3	A carpenter looks for a hammer

7.3 Third Interpretation Experiment

Figure 7.2: The choice options for the four ambiguous items in the third interpretation experiment. On the left the motion event answer is shown, and the right depicts the intensional answer.



Bibliography

[1] Bickerton, D. (1990). *Language and Species.* University of Chicago Press.

[2] Botha, R. (2009). What are windows on language evolution? In: Botha, R. & de Swart, H. (Eds.), *Language evolution: the view from restricted linguistic systems.*, p 1-20. LOT.

[3] Chomsky, N. (1988). *Language and problems of knowledge - the Managua lectures.* Cambridge: MIT Press.

[4] Dryer, M.S. (2008). Order of subject, object and verb. In: Haspelmath, M., Dryer, M.S., Gil, D., & Comrie, B. (eds.), The World Atlas of Language Structures Online. Munich: Max Planck Digital Library, chapter 81. Available online at http://wals.info/feature/81. Accessed on 23-06-2010.

[5] Farkas, D., & de Swart, H. (2003). The semantics of incorporation: from argument structure to discourse transparency. Stanford: CSLI publications.

[6] Forbes, G. (2010). Intensional Transitive Verbs. In: Zalta, E.N. (ed.), The Stanford Encyclopedia of Philosophy (Spring 2010 Edition). Available online at http://plato.stanford.edu/archives/spr2010/entries/intensional-trans-verbs/. Accessed on 26-06-2010.

[7] Gentner, D., & Boroditsky, L. (2009). Early acquisition of nouns and verbs: Evidence from Navajo. In: Mueller Gathercole, V. C. (ed.), *Routes to language: Studies in honor of Melissa Bowerman* (pp. 5-32). Taylor & Francis: New York.

[8] Goldin-Meadow, S., So, W.C., Özyürek, A. & Mylander, C. (2008). The natural order of events: how speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the USA, 105(27)*, p. 9163-9168.

[9] Greenberg, J. H. (1963). *Universals of languages.* Cambridge: MIT Press.

[10] Hauser, M., Chomsky, N., & Fitch, W.T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, p 1569-1579.

[11] Harley, H.B. (2007). Thematic Roles. In: Hogan, P. (ed.), *The Cambridge Encyclopedia of Linguistics*. Cambridge University Press.

[12] Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution.* Oxford University Press.

[13] Johansson, S. (2005). *Origins of language: constraints on hypotheses.* John Benjamins Publishing Company: Amsterdam/Philadelphia.

[14] Kirby, S. (2007). The evolution of language. In: Dunbar, R. and Barrett, L. (eds.), *Oxford Handbook of Evolutionary Psychology*, p 669-681. OUP.

[15] Langus, A., & Nespor, M. (in press). Cognitive systems struggling for word order. *Cognitive Psychology* (2010).

[16] Meir, I., Lifshitz, A., Ilkabasaran, D., & Padden, C. (forthcoming). The interaction of animacy and word order in human languages: a study of strategies in a novel communication task.

[17] Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13(4)*, p 707-784.

[18] Schouwstra, M., Tamariz, M. & Kirby, S. (forthcoming). Semantic structure in a novel communication task.

[19] Spade, P.V. (2008). William of Ockham. In: Zalta, E.N. (ed.), The Stanford Encyclopedia of Philosophy (Fall 2008 Edition). Available online at http://plato.stanford.edu/archives/fall2008/entries/ockham/. Accessed on 20-06-2010.

[20] Wikipedia (2010). Communication. In: Wikipedia, The Free Encyclopedia. Available online at http://en.wikipedia.org/w/index.php?title=Communication&oldid=369893243. Accessed on 27-06-2010.